

**SUBJEKTIVE WAHRSCHEINLICHKEITSTHEORIE
UND RATIONALE ENTSCHEIDUNGSTHEORIE
IN ANWENDUNG AUF DIE RECHTSPRAXIS**

Eleonora Bourmistrov-Jüttner

Disbui. München

1986



SUBJEKTIVE WAHRSCHEINLICHKEITSTHEORIE UND RATIONALE
ENTSCHEIDUNGSTHEORIE IN ANWENDUNG AUF DIE RECHTSPRAXIS

Inaugural-Dissertation
zur Erlangung des Doktorgrades
der Philosophie
an der Ludwig-Maximilians-Universität
zu München

vorgelegt von
Eleonora Bourmistrov-Jüttner
aus München

München 1987

Referent: Prof. Dr. Dr. W. Stegmüller

Korreferent: Prof. Dr. W. Balzer

Tag der mündlichen Prüfung: 28.7.1986.

INHALTSVERZEICHNIS

Einleitende Bemerkungen/Zur Gliederung der Arbeit.....	V
Einleitung.....	VI
I. Moderne Wahrheits- und Erkenntnisproblematik	VI
(1) Das Wahrheitsproblem	VI
(2) Jurisprudenz und Wahrheits- Erkenntnisproblematik.....	X
II. Der Wahrscheinlichkeitsbegriff und die Jurisprudenz.....	XII
III. Rationale Entscheidungstheorie und die Strafrechtspraxis.....	XIII
I. TEIL. SUBJEKTIVE WAHRSCHEINLICHKEITSTHEORIE IN ANWENDUNG AUF DIE STRAFRECHTSPRAXIS	
Einleitung.....	1
Inhaltsüberblick über den I. Teil.....	2
1. Kapitel. Grundlagen der Wahrscheinlichkeitstheorie.....	5
§1. Der Wahrscheinlichkeitsbegriff.....	5
1.1. Der formale Wahrscheinlichkeitsbegriff.....	6
1.2. Verschiedene Interpretationen des Wahrscheinlichkeitsbe- griffs.....	20
1.2.1. Der subjektive Wahrscheinlichkeitsbegriff.....	22
1.2.2. Der objektive (bzw. statistische) Wahrscheinlich- keitsbegriff.....	37
1.2.3. Der logische Wahrscheinlichkeitsbegriff.....	50
§2. Subjektive Wahrscheinlichkeitstheorie und Rechtspra- xis.....	54
2. Kapitel. Anwendung der Wahrscheinlichkeitstheorie auf die Strafrechtspraxis.....	79
Einleitung	79
§1. Diskussion des Wahrscheinlichkeitsbegriffs in der ju- ristischen Literatur.....	84
1.1. Ball	84
1.2. Ekelöf.....	87
1.3. Kingston	92
1.4. Kegel	98
1.5. Schreiber.....	103
1.6. Schilderung des Rechtsfalls 'People v. Collins'	112
1.7. Cullison.....	116

1.8. Finkelstein und Fairley	123
1.9. Tribe	128
1.10. Brilmayer und Kornhauser	139
1.11. Diskussion des Wahrscheinlichkeitsbegriffs in Deutschland..	147
§2. Empirische Untersuchungen.....	150
2.1. Simon und Mahan.....	150
2.2. Marshall und Wise.....	157
2.3. Eggleston.....	162
2.4. Nagel und Neef.....	166
§3. Weitere Diskussion des Wahrscheinlichkeitsbegriffs:Cohen	170
Cohens Paradoxien und ihre Lösung.....	172
§4. Wahrscheinlichkeitstheoretische Rekonstruktion juristi-	
scher Fälle.....	212
 II.TEIL. RATIONALE JURISTISCHE ENTSCHEIDUNGSTHEORIE	
Einleitung.....	233
Inhaltsüberblick über den II.Teil.....	235
1.Kapitel. Grundlagen einer rationalen juristischen Entschei-	
dungstheorie.....	238
§1. Einführung in die rationale Entscheidungstheorie.....	238
§2. Die bisherigen juristischen Entscheidungsmodelle.....	251
2.1. Das konservative Entscheidungsmodell.....	252
2.2. Das Modell von Kaplan.....	253
2.3. Das Modell von Maassen.....	258
2.4. Das Modell von Motsch.....	260
2.5. Das Modell von Marshall&Wise.....	261
2.6. Das Modell von Cullison.....	263
2.7. Das Modell von Nagel&Neef.....	265
§3. Verschiedene entscheidungstheoretische Modelle und die	
Frage ihrer Anwendbarkeit auf die Strafrechtspraxis.....	269
3.1. Die Konzeption von Savage.....	271
3.2. Die Konzeption von Fishburn.....	277
3.3. Die Konzeption von Jeffrey.....	280
2.Kapitel. Rationale juristische Entscheidungsbildung.....	282
§1. Die juristische Entscheidungssituation.....	283
1.1. Intuitive Schilderung des rationalen juristischen Entschei-	
dungsmodells	283
1.2. Quantifizierung des Tatbestands.....	297

§2. Der juristische Nutzen.....	306
2.1. Eine informelle Betrachtung von Zielen.....	307
2.2. Vergleich zwischen verschiedenen Nützlichkeitsmatrizen.....	309
2.3. Formale Darstellung von Zielen.....	312
2.4. Verallgemeinerung verschiedener Zielvorstellungen. Untersu- chung an einem Beispiel	322
2.5. Entscheidungstheorie und einige Gedanken über Ziel und Zweck der Strafe	329
2.6. Die Hierarchie juristischer Ziele.....	337
§3. Axiomatische Festlegung einer rationalen juristischen Entscheidungstheorie.....	340
3.1. Aufhebung der entscheidungstheoretischen Indifferenzbedingung	342
3.2. Beweisgrenzen. Die Rechtfertigung des Beweiskriteriums "mit überwiegender Wahrscheinlichkeit" für die 2x2-Konsequenzenma- trix.....	343
3.3. Das Beweiskriterium "mit kumulativ überwiegender Wahrsein- lichkeit"	350
3.4. Reduktion der entscheidungstheoretischen Beurteilung auf die wahrscheinlichkeitstheoretische aufgrund der Gerechtigkeits- auffassung	363
3.5. Feststellung der Schuld (der subjektive Tatbestand).....	368
3.Kapitel. Kollektive Entscheidungen.....	371
4.Kapitel. Rationale Strafzumessung.....	385
4.1. Das rationale Strafzumessungsmodell von Haag.....	387
4.2. Die Bestimmung der Tatschuld (bzw. der tatschuldangemessenen Strafe) als Messung.....	392
4.3. Ein neues Strafzumessungsmodell.....	397
Anhang.....	407
I.Metrisierungsmethoden	407
I.1. Metrisierung von Nützlichkeiten.....	407
I.2. Metrisierung von Wahrscheinlichkeiten.....	422
II. Die v.Neumann-Morgenstern-Axiome	429
III. Die Theorie von Savage (Axiomatik).....	431
Bibliographie.....	434
Liste der Symbole.....	442

Einleitende Bemerkungen:

In der nachfolgenden Einleitung geht es uns zum einen darum, zu zeigen, daß die moderne Diskussion um die Wahrheitsproblematik für die Jurisprudenz insofern von großer Relevanz ist, als der von der Rechtspraxis erhobene Anspruch auf die Wahrheitsfeststellung bzgl. einer juristischen Sachverhaltshypothese den neuesten Erkenntnissen nicht standhält, und zum anderen darum, die Überleitung zum eigentlichen Gegenstand dieser Arbeit zu geben. Dieser hat allerdings mit der in der Einleitung zu schildernden Wahrheitsproblematik nicht unmittelbar zu tun (auch wenn in der letzteren eine philosophische Ausgangsbasis für die nachfolgenden wissenschaftstheoretischen Überlegungen gesehen werden kann), sondern besteht in der rationalen Rekonstruktion des juristischen Urteils- bzw. Entscheidungsprozesses in einem wahrscheinlichkeitstheoretischen bzw. entscheidungstheoretischen Rahmen.

Zur Gliederung der Arbeit:

Die Arbeit ist entsprechend den beiden Hauptthemen in zwei Teile gegliedert: Im ersten Teil befassen wir uns mit der Präzisierung der Sachverhaltsbeurteilung in der Strafrechtspraxis aufgrund des subjektiven Wahrscheinlichkeitsbegriffs; im zweiten Teil wird eine rationale juristische Entscheidungstheorie aufgestellt. Den beiden Teilen, einzelnen Kapiteln bzw. Paragraphen sind kurze Einleitungen vorangestellt; am Ende jedes Paragraphen (von wenigen Ausnahmen abgesehen) befindet sich jeweils eine kurze Zusammenfassung der zuletzt gewonnenen Ergebnisse und ein Ausblick auf das weitere Vorgehen. Jeder Teil beginnt mit einer kurzen Inhaltsübersicht der dort behandelten Themen. Der zweite Teil kann (von wenigen Stellen abgesehen) unabhängig vom ersten gelesen werden.

EINLEITUNG

I. Moderne Wahrheits- und Erkenntnisproblematik

(1) Das Wahrheitsproblem

Wahrheitsdefinition

Aus philosophischer Sicht definiert man im allgemeinen die Wahrheit als 'Übereinstimmung' von Denken oder Vorstellen mit der Wirklichkeit. Diese Auffassung geht auf Aristoteles zurück. Sie findet jedoch ihren Ausdruck auch in der heutigen formalen Logik.

Kennzeichnend für die moderne Auffassung des Wahrheitsbegriffs, wenn "wahr" in wissenschaftssprachlichen Kontexten gebraucht wird, ist seine explizite Sprachbezogenheit. Danach werden bestimmte Sprechhandlungen, nämlich Behauptungen, als "wahr" ausgezeichnet, wenn sie unabhängig von einer bestimmten Äußerungssituation betrachtet werden können.

In der modernen Auffassung des Wahrheitsbegriffs, die auf Tarski¹⁾ zurückgeht, wird versucht, den aristoteleschen Wahrheitsbegriff in einer neuartigen Weise einzuführen. Dieser Begriff erweist sich vor allem in der mathematischen Grundlagenforschung von außerordentlicher Bedeutung. Auf Tarski geht auch das bekannte Adäquatheitskriterium der Wahrheit zurück: " $\text{wahr}('p') \leftrightarrow p$ ", wobei die Aussage " $\text{wahr}('p')$ " nicht derselben Sprache angehört wie " p ". Dieses Kriterium gilt nur für formale Sprachen, denn nur in diesen ist die Unterscheidung von Sprachebenen explizierbar²⁾.

In der obigen Bestimmung wird " p " durch eine Aussage der Sprache, auf die sich das Wort "wahr" bezieht, ersetzt, und " $'p'$ " durch den Namen dieser Aussage. Z.B. impliziert diese Bestimmung die folgende Teildefinition: "Die Aussage 'Schnee ist weiß' ist wahr genau dann, wenn Schnee weiß ist" (im zweiten Teil dieser Äquivalenz steht eine Aussage, die einen Sachverhalt beschreibt, im ersten der Name dieser Aussage).

1) A.Tarski, Der Wahrheitsbegriff in den formalisierten Sprachen. Deutsche Originalausgabe in: *Studia Philosophica Commentarii Societatis philosophicae Polonorum*. Vol.I, Leopoli (Lemberg) 1935. Abgedruckt in: Berka-Kreiser, Logik-Texte. Kommentierte Auswahl zur Geschichte der modernen Logik. Berlin 1971, S.447-559.

2) Eine weniger formal korrekte, jedoch anschauliche Definition schlägt Tarski auf S.268 (452) vor: "eine wahre Aussage ist eine Aussage, welche besagt, daß die Sachen sich so und so verhalten, und die Sachen verhalten sich eben so und so".

Den Ausgangspunkt der Tarskischen Wahrheitsauffassung bildet das sog. "allgemeine Schema von spezielleren Sätzen" ("Teildefinitionen" bzw. "Erklärungen"):

(*) " x ist eine wahre Aussage dann und nur dann, wenn p " (S.268 (453)).

Diese Formel, die keine Definition der Wahrheit ist, stellt Tarskis berühmte W-Konvention dar.

Das obige Schema erhält man, indem man den beliebigen Einzelnamen einer Aussage mit " x " und die entsprechende Aussage selbst mit " p " angibt. Aus dem Schema erhält man dann "konkrete Erklärungen" bzw. Teildefinitionen, indem man für " p " irgendeine Aussage einsetzt und für " x " einen beliebigen Einzelnamen dieser Aussage. Von diesem Schema aus wird von Tarski die Adäquatheit sowohl des Gebrauchs als auch der Definition des Terms "wahr" festgelegt. Der Term "wahr" wird dann adäquat gebraucht, wenn alle Äquivalenzen der Form (*) gelten. Eine Definition der Wahrheit ist nach Tarski genau dann adäquat, wenn alle diese Äquivalenzen aus ihr folgen.

Eine allgemeine Definition der Wahrheit wäre nach Tarski "gewissermaßen" die logische Konjunktion aller aus dem Schema folgenden Teildefinitionen der Wahrheit. Mehr zu dieser Problematik siehe bei L.B.Puntel, Wahrheitstheorien in der neueren Philosophie (1978); W.Stegmüller, Das Wahrheitsproblem und die Idee der Semantik (1957).

Abgesehen von der oben erwähnten Einschränkung, daß dieses Wahrheitskriterium nur für formale Sprachen gilt, bleibt das fundamentale Problem bestehen, das darin liegt, daß die Übereinstimmung zwischen Aussage und Sachverhalt wiederum Inhalt einer Aussage ist et in infinitum. Außerdem müssen die sogenannten Sachverhalte ihrerseits in Aussagen formuliert werden, so daß die Übereinstimmung genau genommen eine solche zwischen verschiedenen Aussagen ist¹⁾

Neben der Tarskischen semantischen Wahrheitstheorie²⁾, die unter den Sammelbegriff: Korrespondenztheorie³⁾ fällt, gibt es andere Wahrheitstheorien, wie z.B. die rationalistische Wahrheitstheorie (Kohärenztheorie), die in der neueren Zeit vor allem von Neurath und Hempel⁴⁾ vertreten wird. Indem die Kohärenztheorie die Übereinstimmung völlig auf die Sprache reduziert, vermeidet sie die Schwierigkeit, die in der Korrespondenztheorie stets dann auftritt, wenn man sich mit der Frage beschäftigt, zwischen welchen Entitäten die Übereinstimmung genau genommen besteht: Denn nach der Kohärenztheorie ist die Übereinstimmung eine solche zwischen Aussagen. Eine Aussage ist demnach wahr, wenn sie mit den übrigen nicht im Widerspruch steht. Der Nachteil dieser Theorie besteht darin, daß sie kein Kriterium angeben kann, das ein System kohärenter Aussagen vor einem anderen auszeichnet und daher in der reinen Kalkülisierung der empirischen Wissenschaften endet⁵⁾

Die Konsensustheorie (darunter versteht man die pragmatische Wahrheitstheorie (Peirce) und den Konstruktivismus (Habermas)⁶⁾) geht dagegen auf den pragmatischen Ursprung von Aussagen zurück. Danach ist "Wahrheit" ein sozialer Geltungsanspruch, der Aussagen bzw. Regeln zukommt. Unter dem Wahrheitskriterium versteht man im Rahmen dieser Theorie einen Konsens oder eine Übereinstimmung, der die in totalen Sprechsituationen erzielt wird. Eine Aussage ist demnach wahr, wenn alle kompetenten Subjekte der zugrundeliegenden Behauptung hätten zustimmen können. Die Verfahren, mit denen solche Zustimmungen erreicht werden, sind die Begründungsdiskurse.

1) Zur weiteren Kritik am Tarskischen Wahrheitsbegriff vgl. I.M.Bocheński, Die zeitgenössischen Denkmethoden (⁷1975), S.67-72.

2) Die Bezeichnung "semantische Wahrheitstheorie" geht auf Tarski zurück; vgl. Tarski, Die semantische Konzeption der Wahrheit und die Grundlagen der Semantik. In: J.Sinnreich (Hrsg.), Zur Philosophie der idealen Sprache (1972), S.53ff. Abgedruckt in: G.Skirbekk, Wahrheitstheorien. Eine Auswahl aus den Diskussionen über Wahrheit im 20.Jahrhundert (1977).

3) Neben der semantischen Wahrheitstheorie bezeichnet man mit diesem Terminus auch die Abbildungstheorie der Wahrheit und die ontologische Wahrheitstheorie. Gemeinsam ist allen diesen Theorien, daß sie gerade die Übereinstimmung von Aussagen mit einem außersprachlich erfäßbaren Sachverhalt (Gegenstand) als Ausgangspunkt ihrer Wahrheitsdefinition nehmen. Da die semantische Wahrheitstheorie eine Präzisierung der herkömmlichen Korrespondenztheorie darstellt, nimmt sie eine Vorrangstellung gegenüber den übrigen Korrespondenztheorien ein.

4) O.Neurath, Soziologie im Physikalismus. In: 2 Erkenntnis (1931), S.393-431; Protokollsätze, in: 3 Erkenntnis (1932/1933), S.204-214, 346-362; C.G.Hempel, On the Logical Positivists' Theory of Truth, in: 2 Analysis (1934/1935), S.49-59. Weitere Literatur zur Kohärenztheorie: N.Rescher, The Coherence Theory of Truth (1973).

5) Eine weitere Kritik stammt von M.Schlick, Über das Fundament der Erkenntnis, in: 4 Erkenntnis (1934), S.79-99. Schlick ist der Auffassung, die Kohärenztheorie erörtere überhaupt nicht die Übereinstimmung der Erkenntnis mit der Wirklichkeit und verfehle damit die Hauptsache, denn sie begreife das System der Sätze der Wissenschaft gerade nicht als "das, was es eigentlich ist, nämlich (als) ein Mittel, sich in den Tatsachen zurechtzufinden". Zur Kritik an Schlick vgl. Puntelli, S.179ff.

6) J.Habermas, Wahrheitstheorien, in: H.Fahrenbach (Hrsg.), Wirklichkeit und Reflexion (Festschrift W.Schulz) (1973), S.221-265.

Die zentrale Schwierigkeit für diese Theorie liegt darin, eine Bestimmung für die Kompetenz (z.B. Sachverstand, Vorurteilsfreiheit, Emotionsfreiheit, Autoritätsungebundenheit, Aufrichtigkeit) zu finden, die ohne Rückgriff auf die Regelkompetenz für Begründungsdiskurse eingeführt werden kann.

Es ist im Rahmen dieser Einleitung nicht möglich, auf die vielen anderen Wahrheitstheorien, wie z.B. die verschiedenen sprachanalytischen Wahrheitstheorien, die dialogische Wahrheitstheorie etc. einzugehen¹⁾ Erwähnt sei jedoch noch die Redundanztheorie der Wahrheit (sprachanalytische Wahrheitstheorie), die auf das Wort "wahr" völlig verzichten will, d.h. die Redundanztheorie der Wahrheit behauptet folgendes: Sätze, in denen das Wort "wahr" vorkommt, sind in der Regel redundant, d.h. dieses Wort kann ohne Informationsverlust aus ihnen eliminiert werden. Am Beispiel: Statt zu sagen "Es ist wahr, daß es regnet", kann man ohne jeden Informationsverlust einfach sagen: "Es regnet"²⁾

Verifikations- und Falsifikationskriterien (Erkenntnisproblematik)

Handelt es sich bei der Wahrheitsdefinition um eine konstitutionelle Frage, so ist mit dem Wahrheitskriterium bzw. der Wahrheitsfeststellung eine Geltungsfrage verbunden. Wichtig ist dabei nicht nur, daß ein Behauptungssatz wahr ist, sondern, daß er sich als wahr erweist, bzw. wenn er falsch ist, daß er sich als falsch erweist, d.h. widerlegt ist. Das Kriterium der Wahrheit spielt daher vor allem im Verifikationsprozeß bzw. Falsifikationsprozeß eine Rolle.

Das Verifikationskriterium besagt, daß ein Behauptungssatz empirisch verifizierbar (oder aber analytisch) sein muß, um im Sinne eines Abgrenzungskriteriums signifikant zu sein (d.h. eine kognitive Bedeutung zu haben).

Geschichtlich gesehen spielte die Diskussion um die Verifikationsproblematik in der modernen Philosophie immer eine wichtige Rolle. Am Anfang des Logischen Empirismus vertraten manche seine Anhänger wie M.Schlick oder aber auch der frühe Wittgenstein die Meinung, daß der Sinn eines Satzes durch die Methode seiner Verifikation bestimmt ist. Auch war die Ansicht weit verbreitet, daß man bestimmte Sätze, sog. Beobachtungssätze oder Basissätze, auszeichnen kann, die einer Überprüfung im Sinne der Verifikation oder Falsifikation unmittelbar zugänglich sind.

Auch in der heutigen Wissenschaftstheorie ist Verifikationsproblematik von zentraler Bedeutung. Das erkenntnistheoretische Problem der Auszeichnung einer empirischen Basis muß als bis heute ungelöst betrachtet werden. Dieses Problem tritt allerdings hinter dem Versuch zurück, die Beziehung von Basissätzen zu Nicht-Basissätzen, für die auch das Kriterium der empirischen Signifikanz beansprucht wird, zu charakterisieren. Dieses Kriterium verlangt, daß jeder empirisch sinnvolle Satz durch eine endliche Menge von Basissätzen mittelbar verifizierbar sein muß. Das kann jedoch für nicht endliche Allsätze (dazu zählen beispielsweise Naturgesetze) nicht erfüllt werden, d.h. solche Sätze werden in

1) Einen guten Überblick darüber bietet G.Skirbekk (S.VII Fn.2 dieser Arbeit).

2) W.Franzen, Zur Redundanztheorie der Wahrheit. Ein historischer und systematischer Überblick, in: 12 Conceptus (1978), S.54-69; (1979), S.47-62; Zur neueren Wahrheitsdiskussion: Redundanztheorie versus Korrespondenztheorie der Wahrheit, in: 35 Zeitschrift für philosophische Forschung (1981), S.73-90.

Vgl. dazu auch F.P.Ramsey, Facts and Propositions (1927).

der Regel von dem Kriterium als nicht empirisch sinnvoll erfaßt. Als Lösung dieses Problems sind (vor alle von Carnap und Hempel) anstelle der strengen Form des Kriteriums andere Kriterien formuliert worden, die nur eine schwächere Form der Bestätigungsfähigkeit verlangen.

Einen ganz anderen Ansatz in bezug auf die Geltung wissenschaftlicher Hypothesen hat K. Popper in seiner "Logik der Forschung" (1934) gewählt: Nach ihm können Hypothesen niemals durch Erfahrung verifiziert, sondern höchstens falsifiziert werden. Entsprechend zeichnet nicht ein Verifikationskriterium, sondern ein Falsifikationskriterium wissenschaftliche Sätze aus.

Das Falsifikationskriterium besagt, daß ein Behauptungssatz wissenschaftlich nur dann ist, wenn er prinzipiell durch sogenannte Basissätze falsifizierbar ist.

Eine Konsequenz dieses Kriteriums ist: Existenzsätze über unendliche Individuenbereiche werden, entgegen wissenschaftlicher Praxis, generell für unwissenschaftlich erklärt. Man soll dabei nicht aus den Augen verlieren, daß Popper das Falsifikationsprinzip nicht als Signifikanz-, sondern als Abgrenzungskriterium formuliert hat, das den Bereich wissenschaftlicher Sätze von dem nichtwissenschaftlicher, beispielsweise metaphysischer Sätze, abgrenzen soll. Auch in bezug auf Basissätze vertritt Popper eine eigenständige Position: Basissätze, anhand derer man Hypothesen falsifiziert, sind für ihn hypothetisch; sie können durch die Entscheidung, andere Basissätze dem Falsifikationsprozeß zugrunde zu legen, widerrufen werden¹⁾

Das Wahrheitskriterium im vorangegangenen bezog sich primär auf naturwissenschaftliche Erkenntnisse: Erweist sich dabei die Verifikation von Allhypothesen und die Falsifikation von Existenzsätzen über unendliche Individuenbereiche als unmöglich, so hält man gewöhnlich die Verifikation einer singulären Aussage, insofern sie unmittelbares Beobachtungswissen formuliert, und die Falsifikation von Allhypothese für weniger problematisch bzw. für methodologisch gesichert. Auch die schwächere Form des Signifikanzkriteriums, die Bestätigungsfähigkeit, hat wenigstens im Zusammenhang mit der Frage nach der vorläufigen Akzeptierbarkeit der naturwissenschaftlichen Allhypothesen ihren methodologischen Wert.

Wendet man sich von naturwissenschaftlichen zu historischen und psychologischen Erkenntnissen, so bietet sich hier eine andersartige Situation dar. Zwar sind es meistens Tatsachen über Einzelereignisse oder -vorgänge, mit denen man sich in historischen und psychologischen Wissenschaften befaßt, so daß die oben erwähnte Schwierigkeit bzgl. der Verifikation von Allsätzen und der Falsifikation von Existenzsätzen dabei nicht auftritt, die Feststellung der Wahrheit ist hier jedoch viel problematischer als bei den Naturwissenschaften. Denn historische und psychologische Erklärungen²⁾ Beschreibungen und Begründungen haben einen viel zu stark hypothetischen Charakter: Der Historiker hat es meistens mit Ereignissen

1) Eine Konsequenz des Popperschen "Konventionalismus" ist daher, daß Falsifikation nicht Falschheit impliziert. Auch eine richtige Theorie kann relativ auf akzeptierbare Basissätze falsifiziert werden.

2) Es handelt sich hier eher um Erklärbarkeitsbehauptungen als um richtige Erklärungen, denn entweder werden Gesetzhypothesen, die in Prämissen einer Erklärung auftreten müssen, nicht ausdrücklich erwähnt oder fehlen völlig. Daher können solche Erklärbarkeitsbehauptungen nicht verifiziert, sondern bestenfalls induktiv bestätigt werden. Mehr zur Unterscheidung: Erklärung-Erklärbarkeitsbehauptung-(Pseudoerklärung) siehe bei W. Stegmüller, Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie, Bd. I, Teil 3.

nissen und Vorgängen zu tun, an denen er selbst nicht teilgenommen d.h. sie selbst nicht beobachtet hat; der Psychologe mit fremdpsychischen Phänomenen, die er nicht erlebt. Die Verifikation bzw. Falsifikation durch den Augenschein liegt hier also meistens nicht vor. Daher können Argumente - es handelt sich hier vor allem um induktive bzw. induktiv-probabilistische (d.h. Wahrscheinlichkeits-) Überlegungen und Schlüsse -, die der Historiker bzw. der Psychologe für oder gegen eine Hypothese anführt, zwar der letzteren Gewicht verleihen bzw. ihre Glaubwürdigkeit mindern, sie in der Regel jedoch weder verifizieren noch falsifizieren.

(Wir haben früher erwähnt, daß auch im Rahmen der Konsensustheorie ein Wahrheitskriterium aufgestellt wird: Darunter versteht man einen Konsensus, der in totalen Sprechsituationen erzielt wird. Die intersubjektive Nachprüfbarkeit, die das empirische Wahrheitskriterium verlangt, ersetzen hier Begründungsdiskurse. Das Akzeptieren dieses konventionellen Wahrheitskriteriums hängt nun davon ab, ob solchen Begründungsdiskursen eine empirische Informationsbasis zugrundeliegt oder nicht. Im letzteren Fall bedeutet der Konsens eine notwendige Bedingung, um zum Schluß zu kommen. Ansonsten wird man allem Anschein nach den Konsens durch Überprüfung ersetzen wollen).

(2) Jurisprudenz und Wahrheits-Erkenntnisproblematik

Fragt man, welche Wahrheitsdefinition der juristischen Argumentationspraxis angemessen ist, so scheint in bezug auf Sachverhaltsermittlungen die Auffassung, wonach die Wahrheit in Übereinstimmung einer Aussage mit Tatsachen besteht, nahelegend zu sein.

Eine andere Frage ist, ob man die Wahrheit im juristischen Fall feststellen kann. Der Richter, der darüber entscheidet, ob die Hypothese "Der Angeklagte ist der Mörder" wahr ist, hat die Tat selbst nicht beobachtet. Er kann sie also nicht durch Augenschein verifizieren. Aber auch wenn er das könnte, würde seiner Beobachtung in den allermeisten Fällen die Möglichkeit der intersubjektiven Nachprüfbarkeit fehlen, die einem größeren Kreis von Menschen erlaubt, die Hypothese als verifiziert zu akzeptieren. Im juristischen Fall liegt somit, wie wir sehen, die ganze Problematik der historischen Forschung vor: Man kann die Hypothese bestenfalls durch Tatsachen (juristisch gesprochen: Beweismittel) induktiv bestätigen bzw. durch solche ihre Glaubwürdigkeit mindern, sie jedoch in der Regel weder verifizieren noch falsifizieren.

Aber auch dann, wenn der Richter über die Vorsätzlichkeit der Tat, d.h. ob die Tat dem Angeklagten im Sinne der Verantwortlichkeit persönlich vorzuwerfen ist, entscheidet, begibt er sich auf das Feld hypothetischer, meistens psychologischer, Argumentationen oder Überlegungen. In vielen Fällen wird der Richter aus dem sonst normalen Verhalten des Angeklagten auf den Vorsatz schließen. Im Zweifelsfall wird er seine Überzeugung auf ein Sachverständigengutachten stützen. In beiden Situationen liegt also eine induktive Schlußweise vor, die nicht ausreicht, um bestimmte Sachverhalte, hier psychische Zustände, zu verifizieren bzw. zu falsifizieren.

Dehnen wir die Diskussion um das empirische Signifikanzkriterium auf die juristischen Hypothesen aus, so können wir hier also als ein solches weder ein Verifikations- noch Falsifikationsprinzip, sondern nur ein viel schwächeres Kriterium annehmen, das lediglich die Bestätigungsfähigkeit einer juristischen Hypo-



these verlangt. In diesem Sinne können wie nicht von einer juristischen Hypothese behaupten, daß sie "wahr" ist, sondern lediglich, daß sie uns glaubwürdig, wahrscheinlich bzw. plausibel erscheint¹⁾

Für die Rechtspraxis bedeutet das folgendes: Die Forderung, die die Gesetzgebung an den Richter stellt, nämlich, daß er in bezug auf den Sachverhalt in einem konkreten Fall zur vollen Überzeugung kommt, daß der Sachverhalt wahr bzw. falsch ist²⁾, kann der Richter nicht wissenschaftlich-rational erfüllen, er kann also nie, juristisch gesprochen, den "Wahrheitsanspruch durch Wahrheitsfindung" auf rationale Weise "legitimieren".

1) Der Gebrauch des Wortes "wahr" in logischen Kontexten (und mit solchen hat man es auch in der Jurisprudenz zu tun, wenn man juristische Argumentationen auf ihre formale Richtigkeit hin überprüft), ist dagegen unproblematisch. Unter der 'logischen Wahrheit' versteht man Wahrheit "auf Grund der semantischen Regeln allein". Vgl. W. Stegmüller, Hauptströmungen der Gegenwartsphilosophie, Bd. 11, S. 418f: "Wenn man also z.B., um die Wahrheit eines zu einem semantischen System gehörenden Satzes festzustellen, kein Erfahrungswissen heranziehen muß, sondern für dieses Resultat die semantischen Regeln, nach denen S aufgebaut ist, genügen, so ist der Satz logisch wahr".

2) § 261, 1) StPO lautet: "Die forensische Wahrheit, die Grundlage für die Sachentscheidung des Strafgerichts, ist das, wovon der Richter aufgrund der Hauptverhandlung... voll überzeugt ist". Zur 'Überzeugung' schreibt die StPO folgendes: § 261 "2) A. Überzeugung: Es genügt ein nach der Lebenserfahrung ausreichendes Maß an Sicherheit, demgegenüber vernünftige Zweifel nicht mehr aufkommen... eine persönliche Gewißheit des Richters in diesem Sinn... Daher bleibt ein bloß theoretischer Zweifel an der Schuld unberücksichtigt... Der Richter darf aber nicht von seiner Überzeugung ausgehen, soweit ihr zwingende Gesetze der Logik, feststehende Erkenntnisse der Wissenschaft oder dem Zweifel entthobene Tatsachen der Lebenserfahrung widersprechen. Der Beweis muß vielmehr mit lückenlosen, nach vollziehbaren Argumenten geführt sein..., erst auf dieser Grundlage gewinnt der Richter seine "persönliche Gewißheit" (Unterstreichungen von mir). Trotz der löblichen Absicht stellt die Strafprozeßordnung zu hohe Ansprüche an das menschliche Vermögen. Denn die von ihr aufgestellte Forderung wird wohl kaum in Wirklichkeit erfüllbar sein: Weder reicht i.d.R. die Informationsgrundlage aus, um die "Wahrheit" im Prozeß rational zu behaupten, noch handelt es sich dabei um einen formallogischen Beweis. Angesichts der praktischen Unsicherheit (und nicht nur, wie die StPO abschätzend schreibt, "eines bloß theoretischen Zweifels") fördert der § 261 geradezu eine Überzeugungshaltung, die rational nicht nachvollziehbar ist).

Zur Wahrheitsforderung der Gesetzgebung vgl. auch § 226, 1) StPO: "Das Kernstück des Strafprozesses ist die Hauptverhandlung. In ihr soll nach dem mehr summarischen Vor- und Zwischenverfahren die Wahrheit (1 zu § 261) endgültig festgestellt werden" (Unterstreichungen von mir).

II. Der Wahrscheinlichkeitsbegriff und die Jurisprudenz

Der "Wahrheitsanspruch", wie ihn die Gesetzgebung verlangt, kann also rational nicht legitimiert werden. Man kann natürlich versuchen, den Wahrheitsbegriff im juristischen Sprachgebrauch durch andere, der Sachlage adäquate Begriffe wie "Wahrscheinlichkeit", "Glaubwürdigkeit" oder "Bestätigung" zu ersetzen. Dies muß sogar geschehen, wenn man juristische Sachverhaltshypothesen korrekt beurteilen, oder allgemeiner, die Rechtspraxis rational rekonstruieren will (in diesem Sinn erklärt sich die Thematik der vorliegenden Arbeit). Das Gelingen eines solchen Projekts auch in der Praxis hängt davon ab, ob man bereit ist, auf die (rechtspolitisch) starke Forderung nach der "vollen Überzeugung von der Wahrheit" zugunsten der Rationalität und Korrektheit und damit der schwächeren Forderung nach der Bestätigungsfähigkeit bzw. einer wahrscheinlichkeitstheoretischen Beurteilung des juristischen Sachverhalts, zu verzichten.

In der Tat kann man in der letzten Zeit das steigende Interesse am Wahrscheinlichkeitsbegriff seitens der Juristen und die Tendenz beobachten, die wahrscheinlichkeitstheoretischen Argumentationen stärker in die Rechtspraxis einzubeziehen. Auch die Rechtsprechung vermeidet neuerdings die explizite Bezugnahme auf die "vollständige Wahrheitsfindung" und spricht stattdessen von der "an Sicherheit grenzenden Wahrscheinlichkeit" oder von "einem hohen Grad an Wahrscheinlichkeit", den sie allerdings mit der "Wahrheit" gleichsetzt (Wir werden auf diesen Punkt ausführlich im Haupttext eingehen).

Die Ablösung vom Wahrheitsbegriff und die Hinwendung zum Wahrscheinlichkeitsbegriff im juristischen Denken hat allerdings negative Seiten: So ist in der juristischen Fachliteratur zum Wahrscheinlichkeitsbegriff die Tendenz sehr stark spürbar, den juristischen Wahrscheinlichkeitsbegriff objektiv-statistisch zu begreifen. Bei näherer Betrachtung entsteht jedoch der Eindruck, daß die Autoren meistens von vagen und sogar naiven Vorstellungen vom Wahrscheinlichkeitsbegriff ausgehen, ihn häufig mißdeuten oder ihn für eine Trivialität halten, deren nähere Klärung überflüssig ist. In keiner der Arbeiten, die sich mit den juristischen Wahrscheinlichkeitsurteilen befassen (und auf die wir noch später näher eingehen), findet sich eine Explikation des verwendeten Wahrscheinlichkeitsbegriffs und nur in wenigen einige Regeln aus der Wahrscheinlichkeitsrechnung.

Der dem juristischen Rasonieren zugrundeliegende Wahrscheinlichkeitsbegriff muß geklärt und der bekannte mathematische Wahrscheinlichkeitsbegriff auf präzise Weise eingeführt werden, will man feststellen, ob es sich bei dem letzteren um eine Präzisierung des ersteren handelt, bzw. allgemeiner, will man die Relevanz und Adäquatheit von wahrscheinlichkeitstheoretischen Argumentationen für die Rechtspraxis beurteilen.

Aus diesem Grund beginnen wir unsere Ausführungen im ersten Kapitel des I. Teils mit einer formalen Einführung in die elementare Wahrscheinlichkeitstheorie und einem Überblick über verschiedene Wahrscheinlichkeitsinterpretationen; ferner werden wir die Eigenarten des induktiv-juristischen Rasonierens näher untersuchen¹⁾

1) Der Leser, der weniger an Präzision und formaler Klarheit als an einem raschen Überblick interessiert ist, kann eine informelle Darstellung des Wahrscheinlichkeitsbegriffs und eine Rechtfertigung der subjektivistischen Interpretation bei E.L.Nell, Wahrscheinlichkeitsurteile in juristischen Entscheidungen (1983), finden.

III. Rationale Entscheidungstheorie und die Strafrechtspraxis

Nicht nur die Sachverhalts-, sondern auch die Entscheidungsfrage muß geklärt werden, will man den gesamten juristischen Beurteilungs- und Entscheidungsprozeß rational rekonstruieren.

Bereiche, in denen man praktische Entscheidungen in Verbindung mit subjektiven Bewertungen von unsicheren Ereignissen oder Sachverhalten trifft, werden gewöhnlich rational im entscheidungstheoretischen Rahmen rekonstruiert.

Auch in der Strafrechtspraxis geht man von unsicheren Sachverhalten aus und es werden hier letztlich praktische Entscheidungen getroffen. Legt man der Beurteilung juristischer Sachverhaltshypothesen (wie wir es tun werden) den subjektiven Wahrscheinlichkeitsbegriff zugrunde, oder allgemeiner, präzisiert man juristische Fälle im Rahmen der subjektiven Wahrscheinlichkeitstheorie, so ist eine entscheidungstheoretische Rekonstruktion der juristischen Entscheidungspraxis zwingend.

Der Vorteil einer solchen Rekonstruktion ist, daß man auf korrekte, d.h. konsequente und konsistente, Weise von subjektiven Bewertungen der unsicheren Sachverhalte und der Konsequenzen unter Berücksichtigung etwaiger Möglichkeiten einer Fehlentscheidung zu rationalen Entscheidungen gelangt. Dies vermißt man in der herkömmlichen juristischen Entscheidungspraxis: Denn nicht nur wird dort die Wahrheit des Sachverhalts behauptet, sondern - als Folge davon - die Möglichkeit einer Fehlentscheidung negiert bzw. unberücksichtigt gelassen.

Die Verbindung zwischen der wahrscheinlichkeitstheoretischen Beurteilung des juristischen Sachverhalts und der entscheidungstheoretischen Lösung des Falls wird allerdings bis heute in der Fachliteratur wenig beachtet. Gegenüber der Fülle an Beiträgen (allerdings fast ausschließlich in Aufsatzform) zum Wahrscheinlichkeitsbegriff, nehmen sich die wenigen Artikel zur entscheidungstheoretischen Beurteilung juristischer Fälle dürrig aus.

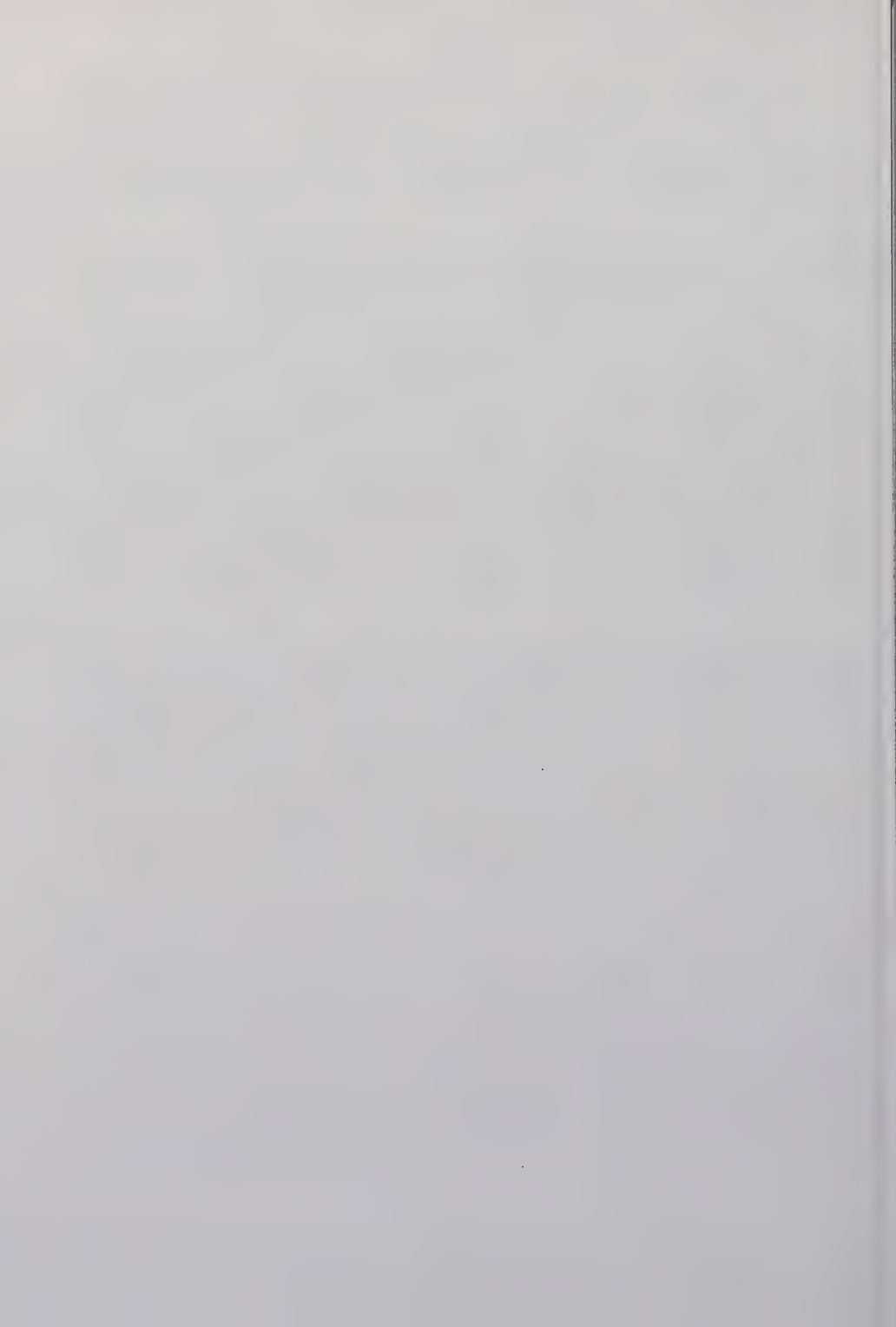
Alle von der juristischen Seite vorgeschlagenen Entscheidungsmodelle leiden an irgendwelchen Mängeln: Entweder sind sie nicht konsistent, nicht adäquat oder einfach nicht ausreichend, um den komplexen juristischen Entscheidungsprozeß wiederzugeben. Auch von der Entscheidungstheorie selbst aus hat man bis heute keinen Versuch unternommen, zu überprüfen, inwiefern die bekannten entscheidungstheoretischen Modelle auch für die Rechtspraxis relevant sind.

Es fehlt also völlig an einer rationalen juristischen Entscheidungstheorie, die nicht nur die Rechtspraxis adäquat wiedergibt, sondern auch den Begriff des juristischen Nutzens klärt.

Die Aufstellung einer rationalen juristischen Entscheidungstheorie ist ein unbestreitbar wichtiges Anliegen, und wir werden im II. Teil dieser Arbeit zeigen, daß die Durchführung dieser Aufgabe gelingen kann. Vorausgreifend sei hier gesagt, daß es sich dabei um ein zweistufiges Entscheidungsmodell handeln wird: Die Wahl eines juristischen Strafrahmens wird als eine solche unter Risiko und die anschließende Wahl einer konkreten Strafe aus dem Strafrahmen als ein Optimierungsproblem unter Sicherheit verstanden. Bei dem Entscheidungsmodell unter Risiko wird es sich um ein Savage-Modell handeln, das wir durch die Hinzunahme einiger zusätzlicher Axiome auf die juristische Entscheidungssituation in geeigneter Weise einschränken.

I. TEIL

SUBJEKTIVE WAHRSCHEINLICHKEITSTHEORIE IN ANWENDUNG AUF DIE
STRAFRECHTSPRAXIS



I. Teil: Subjektive Wahrscheinlichkeitstheorie in Anwendung auf die Strafrechtspraxis

Einleitung

In der allgemeinen Einleitung zu dieser Arbeit sprachen wir davon, daß man den Begriff der Wahrheit durch denjenigen der Wahrscheinlichkeit ersetzen muß, will man eine juristische Sachverhaltshypothese adäquat beurteilen. In diesem Teil geht es uns darum, diese Argumentation weiter zu vertiefen. Es wird gezeigt, daß es sich bei der Rechtspraxis, in welche die Feststellung und Beurteilung juristischer Sachverhaltshypothesen fällt, um einen Bereich des induktiven (intuitiven, alltäglichen) Rasonierens handelt; darüber hinaus, daß induktive juristische Urteile, Überlegungen, Erwägungen und dergl. solche subjektiv-probabilistischer Art sind. Wir wollen allerdings bei dieser Erkenntnis nicht stehenbleiben, sondern werden die Möglichkeit untersuchen, diese (in der traditionellen Rechtspraxis vagen, mehrdeutigen und häufig emotionsbeladenen) Überlegungen mit den dafür geeigneten Mitteln zu präzisieren. Das Hauptanliegen besteht also in der rationalen Rekonstruktion juristischer Sachverhaltsbeurteilungen aufgrund des subjektiven Wahrscheinlichkeitsbegriffs.

Mit einer solchen rationalen Rekonstruktion verbinden wir den Anspruch, Korrektheit, Unparteilichkeit, Gerechtigkeit und intersubjektive Nachvollziehbarkeit der einzelnen Urteilsbildung in der Rechtspraxis besser gewährleisten zu können als es mit den herkömmlichen, traditionellen juristischen Mitteln möglich ist.

Allerdings ist mit der wahrscheinlichkeitstheoretischen Präzisierung der Sachverhaltsbeurteilung bei weitem noch nicht die Rationalisierung des gesamten Rechtsprozesses von der Feststellung des Sachverhalts bis zum Urteilsspruch gewährleistet. Denn es handelt sich bei diesem nicht nur um den Bereich des induktiven Rasonierens, sondern um einen solchen der Entscheidungsbildung schlechthin. Es sind letztlich Entscheidungen, die aufgrund der subjektiven Beurteilung juristischer Sachverhalte gewonnen werden. Auch die juristische Entscheidungsbildung ist präzisierungsbedürftig und wird von uns im II. Teil dieser Arbeit im Rahmen einer rationalen juristischen Entscheidungstheorie präzisiert.

Abschließend sei hier noch kurz darauf hingewiesen, worum es unserer Meinung nach in der Rechtspraxis gehen bzw. nicht gehen soll. Nicht darum, irgendeine "Objektivität" zu erreichen: Denn auch dann, wenn juristische Sachverhaltsurteile präzisiert werden können, bleiben sie immer noch subjektive Wahrscheinlichkeitsurteile, d.h. sie können weder Wahrheit noch Objektivität garantieren. Stattdessen sollte man sich um die Rationalität (Exaktheit, Korrektheit, intersubjektive Nachvollziehbarkeit, Unparteilichkeit) des Ergebnisses, das aufgrund der zur Verfügung stehenden Informationen und der subjektiven Einschätzung gewonnen wird, bemühen. Rationalität ist in unseren Augen somit Paradigma und oberstes Ziel jeglichen juristischen Vorgehens.

Inhaltsüberblick über den I. Teil

Das erste Kapitel trägt einen einführenden Charakter in den Problemkreis: Wahrscheinlichkeitsbegriff und Strafrechtspraxis. In §1 wird das mathematische Modell der Wahrscheinlichkeitstheorie geschildert: Wir führen den Wahrscheinlichkeitsbegriff auf formale Weise ein und stellen die einfachsten Grundlagen der Wahrscheinlichkeitstheorie dar. Es werden dabei die für eine solche Einführung zentralen Begriffe "Menge", "Ereignis", "Ereigniskörper", "Möglichkeits- und Wahrscheinlichkeitsraum" erläutert und die Axiome des quantitativen Wahrscheinlichkeitsmaßes angegeben. Außerdem findet der Leser hier die wichtigsten Formeln, wie die der bedingten Wahrscheinlichkeit, das allgemeine Multiplikationsprinzip, die Bayes-Formel usw. Auf dieser Grundlage werden wir sodann verschiedene Interpretations- und damit Anwendungsmöglichkeiten des Wahrscheinlichkeitsbegriffs angeben. Die drei bekannten Interpretationen sind: die objektivistische, die sich mit der Explikation des statistischen (oder objektiven) Wahrscheinlichkeitsbegriffs befaßt, die induktive, der der logische Wahrscheinlichkeitsbegriff zugrunde liegt und die subjektivistische, deren Grundbegriff der Begriff der subjektiven (oder personellen) Wahrscheinlichkeit ist. (Im Rahmen der Darstellung der subjektivistischen Auffassung werden zwei für die (später noch zu schildernde) Anwendung auf die Rechtspraxis wichtige Formeln: die Jeffrey-Formel und die B_{erw} -Formel (eine Erweiterung der Bayes-Formel) angegeben).

In §2 befassen wir uns mit der Klärung des Verhältnisses von subjektiver Wahrscheinlichkeitstheorie und Rechtspraxis. Es wird der in der Einleitung angesprochene Gedanke weiter vertieft, daß es sich bei dem juristischen Argumentieren um ein induktives und darüber hinaus probabilistisches handelt. Es erweist sich, daß der intuitive Wahrscheinlichkeitsbegriff, der den juristischen Urteilen zugrunde liegt, als ein subjektiver Wahrscheinlichkeitsbegriff präzisiert werden kann. Ferner werden wir auf drei verschiedene Strömungen im juristischen Bereich eingehen und ihre Ursachen zu erforschen versuchen. Diese Strömungen sind: das traditionelle Denken in Wahrheitskategorien, das nicht nur den größten Teil der theoretischen Ansätze kennzeichnet, sondern auch noch immer für die Rechtspraxis bzw. Rechtsprechung maßgebend ist (speziell werden wir uns mit der juristischen Forderung nach der "völligen Überzeugung des Richters vom Bestehen der Wahrheit (des Sachverhalts)" auseinandersetzen und dabei auf die Problematik der daraus entspringenden Nicht-Berücksichtigung des Gesamtdatums näher eingehen), das Bestreben, den juristischen Wahrscheinlichkeitsbegriff als den Begriff der objektiven oder statistischen Wahrscheinlichkeit zu deuten (in theoretischen, v.a. deutschsprachigen Arbeiten zum Wahrscheinlichkeitsbegriff) und der Versuch, den juristischen Wahrscheinlichkeitsbegriff abzugrenzen und ihn sich anders strukturiert vorzustellen als den mathematischen Begriff der Wahrscheinlichkeit (die sog. "induktive Wahrscheinlichkeitstheorie" von Cohen). Schließlich wird die theoretische Möglichkeit, juristischen Urteilen statt eines Wahrscheinlichkeitsbegriffs einen solchen des "Tatsachengewichts" ("weight of evidence" im Sinne von Keynes) zugrunde zu legen, untersucht.

Im zweiten Kapitel befassen wir uns mit der eigentlichen Anwendung der Wahrscheinlichkeitstheorie auf die Strafrechtspraxis.

§1 faßt die wesentlichen theoretischen Ergebnisse aus der Diskussion des Wahrscheinlichkeitsbegriffs bzw. der Anwendung der Wahrscheinlichkeitstheorie auf die Rechtspraxis in der juristischen Fachliteratur zusammen, wobei im Vordergrund die vorgeschlagenen wahrscheinlichkeitstheoretischen und statistischen Methoden wie auch deren kritische Bewertung stehen.

Der Leser wird mit den für die Behandlung juristischer Sachverhaltshypothesen wichtigsten Techniken bekanntgemacht. Daneben werden aber auch solche Ansätze diskutiert, die zwar in der juristischen Fachliteratur noch immer akzeptiert werden, sich jedoch bei näherem Hinsehen als nicht korrekt bzw. nicht adäquat herausstellen. Wir diskutieren außerdem zwei Arbeiten, die programmatisch für die Haltung derjenigen Juristen stehen können, die die Anwendung wahrscheinlichkeitstheoretischer Methoden in der Rechtspraxis ablehnen. Es erweist sich, daß viele Einwände dieser Autoren aus falschverstandenen konzeptionellen bzw. epistemologischen oder einfach antirationalistischen Gründen erhoben wurden.

Im Unterschied zu diesem theoretischen Paragraphen trägt die Diskussion des Wahrscheinlichkeitsbegriffs im §2 einen empirischen Charakter. Es wird anhand der Ergebnisse mehrerer experimenteller Forschungsarbeiten der Frage nachgegangen, inwiefern das Denken in Wahrscheinlichkeiten auch praktisch für die Rechtspraxis von Bedeutung sein kann. Es wird u.a. die in dieser Arbeit vertretene Auffassung empirisch bestätigt, daß es sich bei der juristischen Überzeugungsbildung um eine solche in subjektiven Wahrscheinlichkeitsurteilen handelt. Außerdem stellt sich heraus, daß manche traditionelle Vorstellung über die Urteilsbildung von mit Wahrscheinlichkeitsurteilen praktisch arbeitenden Juristen korrigiert werden muß, u.a. daß Individuen mit Wahrscheinlichkeiten nicht umgehen können oder daß die intuitive Verarbeitung von Informationen zum Gesamtbild im juristischen Prozeß etwa mit dem Bayesianistischen Ansatz nicht vereinbar sei. Auch der Begriff der "an Sicherheit grenzenden Wahrscheinlichkeit" erfährt eine - von der theoretischen völlig abweichende - empirische Deutung.

Wie bereits erwähnt, vertritt Cohen die Auffassung, daß der juristische Wahrscheinlichkeitsbegriff anders strukturiert ist als der gewöhnliche. Nach Cohen gehorcht er ganz anderen Prinzipien als den Kolmogoroff-Axiomen. Seine Kritik stützt sich darauf, daß der mathematische Wahrscheinlichkeitsbegriff, angewandt auf die Rechtspraxis, dort angeblich zu Paradoxien führt und daher der Argumentation in eben dieser Rechtspraxis nicht angemessen ist.

In §3 werden diese Paradoxien untersucht. Es stellt sich heraus, daß diese Paradoxien keine echten Paradoxien sind, denn sie lassen sich relativ leicht lösen. Somit scheint der Kritik derjenigen, die wie Cohen einen besonderen juristischen Wahrscheinlichkeitsbegriff behaupten und ihn unter Berufung auf die Schwierigkeiten des herkömmlichen Wahrscheinlichkeitsbegriffs verteidigen, die Grundlage entzogen zu sein.

Und schließlich befassen wir uns in §4 mit der rationalen Rekonstruktion juristischer Beweise. Zwei juristische Fälle werden dabei wahrscheinlichkeitstheoretisch rekonstruiert. Bei dieser Rekonstruktion wenden wir folgende Formeln an: die Bayes-, die B_{erw} -, die

Jeffrey-Formel und das Multiplikationsprinzip. Es werden außerdem für eine rationale Rekonstruktion der Rechtspraxis wichtige Forderungen wie die des zulässigen, die des relevanten, die des gesamten Datums usw. formuliert, der Begriff der Relevanz in Anlehnung an Gärdenfors neu definiert und Begriffe wie "Apriori-Wahrscheinlichkeit", "Zeugenglaubwürdigkeit" etc. erläutert.

1. Kapitel: Grundlagen der Wahrscheinlichkeitstheorie

§1. Der Wahrscheinlichkeitsbegriff

In diesem Paragraphen werden der Wahrscheinlichkeitsbegriff eingeführt und die einfachsten Grundlagen der Wahrscheinlichkeitstheorie dargestellt. Als Ziel verfolgen wir dabei eine vollständige und gründliche Darstellung des Wahrscheinlichkeitsbegriffs, soweit dies im Rahmen einer für Juristen gedachten Einführung in die Wahrscheinlichkeitstheorie möglich ist.

Wie bereits in der Einleitung erwähnt,¹⁾ entsteht beim Lesen von Texten, die sich mit der Verwendung des Wahrscheinlichkeitsbegriffs in der juristischen Praxis befassen, leicht der Eindruck, daß die Verfasser entweder nicht wissen, wovon sie reden, oder den Wahrscheinlichkeitsbegriff für eine Trivialität halten, dessen konventionelle vorwissenschaftliche Bedeutung für juristische Zwecke völlig ausreiche, und daher keine Mühe auf sich nehmen, diesen Begriff zu präzisieren. Ohne Klarheit in bezug auf den Wahrscheinlichkeitsbegriff - seine Struktur und seine Interpretation - ist es aber unmöglich, die Relevanz und Adäquatheit der Wahrscheinlichkeitstheorie für die Rechtspraxis zu beurteilen. Dies endlich nachzuholen, ist daher unumgänglich.

Unsere Intention dürfte somit klar sein: Wir wollen bei der Darstellung der Grundlagen der Wahrscheinlichkeitstheorie systematisch vorgehen. Konkret heißt das, daß wir zuerst die Struktur des Wahrscheinlichkeitsbegriffs erläutern (Abschnitt 1.1) und dann auf dieser Grundlage verschiedene Deutungs- und damit Anwendungsmöglichkeiten des Wahrscheinlichkeitsbegriffs (Abschnitt 1.2) angeben. Dem nicht versierten Leser kann der Abschnitt 1.1 trotz seines einführenden Charakters ziemliche Schwierigkeiten bereiten. Er muß daher diesen Abschnitt nicht vollständig bei der ersten Lektüre lesen. Es ist jedoch empfehlenswert, sich folgende Stellen anzuschauen:

- (1) Definition des Ereignisses und der Menge, S.7.
- (2) Definition des Möglichkeitsraumes, des Ereigniskörpers und des Wahrscheinlichkeitsmaßes, S.11,13,14.
- (3) die Axiome des quantitativen endlich additiven Wahrscheinlichkeitsraums, S.15.
- (4) bedingte Wahrscheinlichkeiten, das Multiplikationsprinzip, das allgemeine Additionsprinzip, S.15ff.
- (5) die Bayes-Formel, S.18.

und außerdem alle nicht formal geschriebenen Passagen.

1) vgl. S.XII dieser Arbeit.

1.1. Der formale Wahrscheinlichkeitsbegriff

In diesem Abschnitt¹⁾ führen wir den Begriff der Wahrscheinlichkeit als einen undefinierten Begriff ein und charakterisieren ihn nur axiomatisch. Es wird somit hier zunächst von den möglichen Deutungen des Wahrscheinlichkeitsbegriffs mit dem Ziel abstrahiert, Eigenschaften dieses Begriffs angeben zu können, die unabhängig davon gelten, wie man ihn interpretiert.

Ein solcher undefinierter, axiomatisch charakterisierter Wahrscheinlichkeitsbegriff heißt formaler oder mathematischer Wahrscheinlichkeitsbegriff, und die Theorie, der dieser Begriff zugrunde liegt, mathematische Wahrscheinlichkeitstheorie oder Wahrscheinlichkeitsrechnung (= Wahrscheinlichkeitskalkül). Die Wahrscheinlichkeitsrechnung lehrt, wie aus gegebenen Wahrscheinlichkeiten weitere Wahrscheinlichkeiten zu gewinnen sind. Dagegen sagt sie nichts darüber aus, was "Wahrscheinlichkeit" ist. Sie gibt nur eine notwendige Bedingung für die korrekte Interpretation dieses Ausdrucks an: Nur etwas, was bestimmten, in den Kolmogoroff-Axiomen²⁾ festgehaltenen Merkmalen genügt, darf Wahrscheinlichkeit genannt werden³⁾. Der Wahrscheinlichkeitsbegriff wird im axiomatischen Kalkül nur durch diese formalen Merkmale charakterisiert. Diese Einschränkung - in Form der Forderung, daß der Wahrscheinlichkeitsbegriff die Kolmogoroff-Axiome erfüllt - gilt für alle möglichen Deutungen von "Wahrscheinlichkeit".

Kurze Schilderung der in diesem Abschnitt gestellten Aufgabe und der Vorgehensweise: Unser Ziel ist es, den Begriff des "Wahrscheinlichkeitsmaßes" auf systematische Weise einzuführen. Diesen Begriff können wir als Bestandteil eines umfassenderen mathematischen Modells, des sog. "Wahrscheinlichkeitsraumes" definieren. In diesem Modell kommen außerdem zwei andere für die Wahrscheinlichkeitstheorie zentrale Begriffe vor: der Begriff des "Möglichkeitsraumes" und der Begriff des "Ereigniskörpers über diesem Möglichkeitsraum". Um alle diese Begriffe einführen zu können, bedarf es eines mathematischen bzw. mengentheoretischen Begriffsapparates. So werden wir die Begriffe "Menge", "mengentheoretische Operationen", "Relation", "Funktion" usw. definieren müssen, bevor wir das Instrumentarium parat haben, um auf einwandfreie Weise auch den Begriff "Wahrscheinlichkeitsmaß" definieren und ihn durch Axiome charakterisieren zu können. Anschließend werden wir die wichtigsten wahrscheinlichkeitstheoretischen Formeln, darunter die Formel der bedingten Wahrscheinlichkeit, das allgemeine Multiplikationsprinzip und die Bayes-Formel, angeben.

1) Diesem Abschnitt liegt v.a. die Darstellung von W. Stegmüller, Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie, Bd. IV, Personelle und Statistische Wahrscheinlichkeit (1973), Teil A, S. 107ff, zugrunde.

2) Die Bezeichnung "Kolmogoroff-Axiome" geht auf den russischen Wahrscheinlichkeitstheoretiker A.N. Kolmogoroff zurück, der diese Axiome erstmals präzise formuliert hat.

3) Vgl. Stegmüller, *ibid*, S. 109.

Beschreibung des mathematischen Modells der Wahrscheinlichkeitstheorie

Grundbegriffe

Ereignisse sind jene Entitäten, denen man Wahrscheinlichkeiten zuordnet. Diese Ereignisse werden im Rahmen der mathematischen Theorie als Mengen bestimmter Art konstruiert¹⁾.

Mengen sind abstrakte Zusammenfassungen von Objekten, die als Elemente der Menge bezeichnet werden.

Die in der Mathematik übliche Bezeichnung für die Menge ist das Symbol " $\{ \}$ ", wobei die Elemente der Menge, durch Kommas getrennt, zwischen den beiden geschweiften Klammern angeführt werden²⁾.

Will man von der Zugehörigkeit eines Objekts zu einer Menge sprechen, so verwendet man das Symbol " $\in$ " (" $\in$ " steht für die Elementschiffsrelation). So bedeutet " $x \in A$ " eine Abkürzung für die Aussage "das Objekt x ist Element der Menge A ". Bei endlichen Mengen kann man anstelle der Schreibweise " $x \in \{a, b, \dots, k\}$ " auch die Schreibweise " $x = a \vee x = b \vee \dots \vee x = k$ " verwenden (" $\vee$ " ist das logische Zeichen für "oder")³⁾.

Das definierende Prädikat "F" erlaubt uns, Mengen, deren Elemente ein gemeinsames Merkmal besitzen, zu benennen, ohne die Elemente einzeln anführen zu müssen; denn dieses Prädikat trifft genau auf diese Elemente zu. Mit Hilfe des Prädikats wird die Aussageform " Fx " gebildet und die Menge mit der Menge derjenigen Objekte identifiziert, welche diese Aussageform erfüllen. Die Bezeichnung dafür ist $\{x|Fx\}$; dies ist die Abkürzung für die Aussage "die Menge aller Objekte x , so daß x ein F ist".

" $\{x|Fx\}$ " heißt der der Aussage " Fx " entsprechende Abstraktionsterm; " Fx " heißt die definierende Aussageform der Menge, die dieser Abstraktionsterm bezeichnet.

1) In der subjektivistischen Wahrscheinlichkeitstheorie werden Ereignisse durch Sätze oder Propositionen dargestellt. Erstmals gelang es jedoch Carnap, Propositionen als Mengen zu deuten.

2) Dabei kommt es nicht auf die Reihenfolge der Elemente an. So kann man die Menge $\{a, b, c\}$, d.h. eine Menge, die aus drei Elementen a, b und c besteht, auf sechs verschiedene Weisen schreiben, z.B. als $\{b, a, c\}$, $\{c, b, a\}$ usw.

3) Wir werden außer $\vee$ für "oder" folgende logische Symbole als logische Operatoren verwenden: $\wedge$ für "und", $\neg$ für Negation, $\forall$ für "für alle gilt, ...", $\exists$ für "es gibt, ...", $\rightarrow$ für "wenn..., dann..." und $\leftrightarrow$ für "...genau dann, wenn..."

Die eben eingeführte Symbolik erlaubt uns, Operationen mit Mengen auf die logischen Operationen zurückzuführen. Dies ist insofern wichtig, als dadurch die für Mengenoperationen geltenden Regeln auf die entsprechenden Regeln für logische Zeichen gegründet werden können. Sei a die Abkürzung für die Menge $\{x|Fx\}$. Dann kann man die Komplementärmenge $\bar{a}$ ¹⁾ als eine Abkürzung für die Menge $\{x|\neg Fx\}$ ²⁾ darstellen. Wenn $a = \{x|Fx\}$ und $b = \{x|Gx\}$, so ist der Durchschnitt $a \cap b$ die Menge $\{x|Fx \wedge Gx\}$ und die Vereinigung $a \cup b$ die Menge $\{x|Fx \vee Gx\}$. Man kann somit die Operationen der Bildung des Komplements, des Durchschnitts und der Vereinigung auf die logischen Operationen der Bildung der Negation, der Konjunktion und der Adjunktion zurückführen. Die Gültigkeit der elementaren Regeln der Mengenoperationen ist eine unmittelbare Folge daraus.

Seien a, b und c Mengen. Es gelten folgende Gesetze:

- (1) $a \cap a = a$; (2) $a \cup a = a$; (3) $a \cap b = b \cap a$; (4) $a \cup b = b \cup a$;
 (5) $a \cup (b \cap c) = (a \cup b) \cap c$; (6) $a \cap (b \cup c) = (a \cap b) \cup c$; (7) $(a \cap b) \cup a = a$;
 (8) $(a \cup b) \cap a = a$; (9) $a \cap (b \cup c) = (a \cap b) \cup (a \cap c)$;
 (10) $a \cup (b \cap c) = (a \cup b) \cap (a \cup c)$.

Zwei besondere Mengen sind die Allmenge V , zu deren Elementen sämtliche Objekte des Bereichs gehören, und die leere Menge $\emptyset$, die kein einziges Objekt als Element enthält. Es gelten folgende Definitionen:

$$V := \{x|x = x\} \quad \text{und} \quad \emptyset := \{x|x \neq x\}.$$

Außerdem gelten folgende Gesetze:

- (11) $a \cap \bar{a} = \emptyset$; (12) $a \cap \emptyset = \emptyset$; (13) $a \cup \bar{a} = V$; (14) $a \cup V = V$.

Die Gesetze (1)-(8) besagen, daß die Mengen in bezug auf die beiden Operationen $\cap$ und $\cup$ einen Verband bilden. Zusammen mit den Gesetzen (9) und (10) stellen die Mengen bzgl. dieser beiden Operationen einen distributiven Verband dar. Mit den restlichen vier Gesetzen (11)-(14) ergibt sich, daß die Mengen in bezug auf die drei Operationen der Durchschnitts-, Vereinigungs- und Komplementbildung einen sog. komplementären distributiven Verband bilden, auch Boolescher Verband oder Boolesche Algebra genannt.

Folgende Definitionen erweisen sich von Nutzen:

$a - b := a \cap \bar{b}$; die mengentheoretische Differenz. Diese Differenz soll die Menge derjenigen Objekte bezeichnen, die in a , jedoch nicht in b liegen³⁾.

$a \subseteq b := a \cap b$; die Relation des Mengeneinschlusses; sie kann per definitionem auf den Mengendurchschnitt zurückgeführt werden. Eine andere (äquivalente) Definition für den Mengeneinschluß:
 $a \subseteq b := \Lambda x(x \in a \rightarrow x \in b)$.

Will man Vereinigungs- und Durchschnittsbildung auf Mengen beliebig vieler Mengen erweitern, so ist es zweckmäßig, die Mengen zu nummerieren; dies erreicht man durch das Kennzeichnen von Mengen mit unteren Indizes. Die Indizes bilden ihrerseits eine Zahlenmenge J ,

1) Die Komplementbildung wird stets durch den horizontalen Strich über einem Mengensymbol dargestellt.

2) Dies ist die Bezeichnung für die Menge der Objekte, die nicht die Eigenschaft F besitzen.

3) Eine andere Bezeichnung: $a \setminus b$ (zu lesen: "a ohne b").

Indexmenge genannt.

Sei a_i eine Menge. Dann wird die Menge von a_i , $i \in J$, abkürzend mit $\{a_i\}_{i \in J}$ bezeichnet. Es gilt folgende Definition:

$$\{a_i\}_{i \in J} := \{b \mid \forall i (i \in J \wedge b = a_i)\}.$$

Damit wird die Menge aller derjenigen Mengen b bezeichnet, welche für einen geeigneten Index i mit a_i identisch sind.

Sei $K = \{a_i\}_{i \in J}$. Der Durchschnitt $\bigcap K$ und die Vereinigung $\bigcup K$ sind dann wie folgt festgelegt:

$$\bigcap K := \{x \mid \bigwedge b (b \in K \rightarrow x \in b)\}$$

d.h. $\bigcap K$ ist die Menge aller Objekte, die Elemente sämtlicher Mengen von K sind.

$$\bigcup K := \{x \mid \bigvee b (b \in K \wedge x \in b)\}$$

d.h. $\bigcup K$ ist die Menge aller Objekte, die in mindestens einer Menge von K enthalten sind.

$$\langle a, b \rangle := \{\{a\}, \{a, b\}\}$$

$\langle a, b \rangle$ heißt geordnetes Paar; es gilt: $\langle a, b \rangle \neq \langle b, a \rangle$.

$$a \times b := \{z \mid \forall x \forall y (z = \langle x, y \rangle \wedge x \in a \wedge y \in b)\}$$

$a \times b$ heißt das Cartesische Produkt¹⁾ zweier Mengen a und b .

Zweistellige Relationen können als Mengen geordneter Paare dargestellt werden.

r ist eine zweistellige Relation genau dann, wenn gilt:

$$r \subseteq a \times b.$$

(Nach dieser Definition kann man eine zweistellige Relation als eine Teilmenge des Cartesischen Produkts zweier Mengen darstellen).

Sei R die Bezeichnung für die Menge dieser Relationen, so lautet die Definition dieses Begriffs wie folgt:

$$R := \{r \mid \forall a \forall b (r \subseteq a \times b)\}^{2)}$$

1) Eine andere Bezeichnung dafür ist "Kreuzprodukt".

2) Statt als Mengennamen können wir " R " als Prädikat und daher " $R(x)$ " als Aussageform auffassen. Dann können wir den Begriff der Relation auch wie folgt einführen:

$$R(r) := \forall a \forall b (r \subseteq a \times b).$$

Die Menge der Relationen wird dann nicht mehr durch R , sondern durch $\{r \mid R(r)\}$ bezeichnet.

Die obige Definition ist wegen der sog. Abstraktionsregel möglich, d.h. es gilt nach dieser Regel $y \in \{x \mid Fx\} \leftrightarrow Fy$. Das bedeutet, daß ein Objekt genau dann ein Element der mittels einer Aussageform " Fx " definierten Menge ist, wenn dieses Objekt die Aussageform erfüllt.

Um den Begriff der Funktion definieren zu können, brauchen wir außer dem Begriff der Relation den der Rechtseindeutigkeit und denjenigen der Linkstotalität:

$$\text{Re}(s) := \leftrightarrow \bigwedge x \bigwedge y \bigwedge z ((\langle x, y \rangle \in s \wedge \langle x, z \rangle \in s) \rightarrow y = z),$$

d.h. falls s eine Relation ist, dann besagt die Aussage $\text{Re}(s)$, daß zwei zu s gehörende Paare mit identischem Erstglied auch dasselbe Zweitglied besitzen (Rechtseindeutigkeit).

$$\text{Lt}(s) := \leftrightarrow \bigwedge x \bigvee y (\langle x, y \rangle \in s),$$

d.h. falls s eine Relation ist, dann besagt die Aussage $\text{Lt}(s)$, daß zu jedem Erstglied ein Zweitglied existiert, so daß das geordnete Paar zu s gehört (Linkstotalität).

Der Begriff der Funktion ("Fkt") läßt sich somit definieren als

$$\text{Fkt}(f) := \leftrightarrow R(f) \wedge \text{Re}(f) \wedge \text{Lt}(f).$$

Wir geben hier noch die zwei aus mathematischer Sicht wichtigen Begriffe des Definitions- und des Bildbereiches an.

$$D_I(s) := \{x \mid \bigvee y (\langle x, y \rangle \in s)\}, \quad D_I(s) \text{ ist Vorbereich von } s.$$

$$D_{II}(s) := \{y \mid \bigvee x (\langle x, y \rangle \in s)\}, \quad D_{II}(s) \text{ ist Nachbereich von } s.$$

Wenn f eine Funktion ist, so bezeichnet man $D_I(f)$ auch als den Definitionsbereich oder Argumentbereich (aber auch als Urbildbereich) und $D_{II}(f)$ als den Wertebereich bzw. Bildbereich von f . Die Elemente dieser beiden Bereiche werden Argumente und Werte (bzw. Bildelemente) genannt. Ist eine Funktion f mit $D_I(f) = a$ gegeben, so sagt man auch, f sei als Funktion auf a definiert.

Gegeben sei ein Bereich Q von Objekten sowie eine Funktion f , deren Argumente Teilmengen¹⁾ von Q sind; d.h. $D_I(f)$ soll eine Menge von Mengen von Objekten aus Q sein. Die Funktion, deren Argumentbereich Teilmengen von Q enthält, heißt Mengenfunktion.

Es kann der Fall eintreten, daß alle Teilmengen von Q mögliche Argumente von f sind. Die Menge aller dieser Teilmengen heißt die Potenzmenge $\text{Pot}(Q)$ von Q . Sie ist definiert als:

$$\text{Pot}(Q) := \{a \mid a \subseteq Q\}.$$

Im obigen Fall gilt somit: $D_I(f) = \text{Pot}(Q)$.

Wir wollen hier noch kurz den Begriff des Funktionswertes erläutern: Wenn f eine Funktion ist und x ein Element des Definitionsbereichs von f , also $x \in D_I(f)$, so verstehen wir unter dem Funktionswert $f(x)$

1) a ist eine Teilmenge von b genau dann, wenn gilt: $a \subseteq b$; vgl. die Definition des Mengeneinschlusses, S.8.

das nach Definition eindeutig bestimmte Objekt $y \in D_{\Pi}(f)$, so daß gilt: $\langle x, y \rangle \in f$.

Mit den obigen Begriffen und Definitionen haben wir nun ein Instrumentarium, mittels dessen wir den Begriff "Wahrscheinlichkeitsmaß", der eine Mengenfunktion ist, einführen können.

Die zuletzt eingeführten Begriffe sollen deshalb wahrscheinlichkeitstheoretisch gedeutet werden.

Wahrscheinlichkeitstheoretisch gesehen ist der Grundbereich Q der Raum der möglichen Resultate, kurz "Möglichkeitsraum" genannt, d.h. die endlich oder unendlich vielen Elemente von Q sind sog. mögliche Resultate oder Ergebnisse¹⁾. Q wird als die Menge aller möglichen Resultate aufgefaßt.

Die Ereignisse, die man wahrscheinlichkeitstheoretisch beurteilt, werden als Teilmengen von Q (die genau die Resultate enthalten, in denen die Ereignisse gelten) konstruiert. Die Wahrscheinlichkeiten für das Eintreten von Ereignissen werden durch eine Funktion beschrieben, die jedem möglichen Ereignis die Wahrscheinlichkeit seines Eintretens als numerischen Wert zuordnet. Diese Funktion wird Wahrscheinlichkeitsmaß genannt.

Die Begriffe "Möglichkeitsraum", "mögliche Resultate" und "Ereignis" sollen an einem Beispiel verdeutlicht werden.

Angenommen, Q (= Möglichkeitsraum) besteht aus den sechs möglichen Würfergebnissen (= mögliche Resultate) eines Würfels. Dann kann man nicht nur den Ereignissen "eine 1 wird geworfen" oder "eine 2 wird geworfen" usw. Wahrscheinlichkeiten zuordnen, sondern auch den Ereignissen "eine gerade Zahl wird geworfen", "keine 3 wird geworfen" usw. Damit dürfte wohl klar sein, warum Ereignisse als Teilmengen und nicht als Elemente von Q konstruiert werden. Auch die Elementarereignisse, wie "eine 1 wird geworfen", "eine 2 wird geworfen" usw. sind Teilmengen der Menge Q ; wir ordnen somit Wahrscheinlichkeiten in diesen Fällen nicht den möglichen Resultaten, sondern deren Einermengen zu.

1) Die Elemente von Q können aber auch - wie es in der subjektivistischen Wahrscheinlichkeitstheorie der Fall ist - als mögliche Zustände (der Welt) gedeutet werden. Wir wollen uns hier jedoch der Einfachheit halber an die Bezeichnung der Elemente von Q als mögliche Resultate halten. Erst später - bei der Einführung in die rationale Entscheidungstheorie (II. Teil) - wird die Bezeichnung "mögliche Zustände (bzw. Umstände)" wichtig.

Der Begriff des Wahrscheinlichkeitsraumes

Körper und σ -Körper von Ereignissen

Wie oben festgestellt, bildet die Gesamtheit aller möglichen Resultate den Möglichkeitsraum Q . Ereignisse, denen man Wahrscheinlichkeiten zuordnet, werden als Teilmengen von Q aufgefaßt. Die Menge der Ereignisse kann mit der Potenzmenge $\text{Pot}(Q)$ ¹⁾ identifiziert werden, d.h. mit der Menge aller Teilmengen des Möglichkeitsraumes, Q selbst (Q wird auch als Ereignis gedeutet und das sichere Ereignis genannt) und die leere Menge $\emptyset$ (das Komplement von Q , auch das unmögliche Ereignis²⁾ genannt) eingeschlossen³⁾.

Die Menge der Ereignisse -- unabhängig davon, ob sie mit $\text{Pot}(Q)$ identifiziert wird oder nicht - muß immer die folgende Forderung erfüllen:

Die Menge der Ereignisse soll das sichere Ereignis Q enthalten; ferner soll sie zu jedem Ereignis E auch das Komplement $\bar{E}$ sowie zu je zwei Ereignissen E_i und E_j auch deren Vereinigung $E_i \cup E_j$ enthalten⁴⁾.

Eine Menge von Ereignissen, die dieser Forderung genügt, wird ein Mengenkörper oder eine Mengenalgebra genannt. Es gilt die folgende Definition:

1) Im diskreten Fall, d.h. wenn Q abzählbar (endlich oder höchstens abzählbar unendlich) ist, wird meistens als Definitionsbereich eines Wahrscheinlichkeitsmaßes die Potenzmenge von Q gewählt. Im nicht-abzählbaren Fall ist es dagegen unmöglich, die Potenzmenge von Q zu wählen. Letzteres braucht uns aber hier nicht weiter zu interessieren, denn für unsere Zwecke genügt die Schilderung des diskreten Falls völlig.

2) Es soll nach Voraussetzung immer ein Ereignis aus Q realisiert werden; $\emptyset$ steht somit für das Ereignis, daß kein Ereignis aus Q realisiert ist.

3) Im Beispiel auf der letzten Seite besteht die Potenzmenge aus 36 Elementen. Eine viel einfachere Potenzmenge ergeben die Würfe mit einer Münze. Diese ist $\{\{\text{Kopf}\}, \{\text{Zahl}\}, \{\text{Kopf}, \text{Zahl}\}, \emptyset\}$.

4) Wenn man statt von Ereignissen von Aussagen ausgeht, die diese Ereignisse beschreiben - wie das in der subjektivistischen Wahrscheinlichkeitstheorie der Fall ist - so lautet diese Forderung wie folgt: Die Menge der Aussagen soll die tautologische Aussage, d.h. die Aussage, die das sichere Ereignis beschreibt, enthalten und mit jeder Aussage auch deren Negation sowie zu zwei beliebigen Aussagen auch deren Adjunktion.

D1 Eine Menge Ω von Teilmengen einer gegebenen Menge Q heißt Mengenkörper über Q genau dann, wenn gilt:

- (1) $Q \in \Omega$;
- (2) für jedes $E \in \Omega$ ist auch $\bar{E} \in \Omega$;
- (3) wenn $E_i, E_j \in \Omega$, dann auch $E_i \cup E_j \in \Omega$.

D.h. ein Mengenkörper über Q ist eine Menge, die Q enthält und bezüglich der beiden mengentheoretischen Operationen der Komplementbildung und der Vereinigungsbildung abgeschlossen ist¹⁾.

In der Wahrscheinlichkeitstheorie verwendet man neben der Bezeichnung "Mengkörper" auch die Bezeichnung "Ereigniskörper über dem Möglichkeitsraum" oder einfach "Ereigniskörper".

Anstatt D1 können wir aber auch eine stärkere Definition angeben. Nach dieser Definition wird nicht nur verlangt, daß zu je zwei Mengen auch deren Vereinigung ein Element von Ω sein soll, sondern daß für jede Folge $E_1, E_2, \dots, E_n, \dots$ von Mengen auch deren unendliche Vereinigung dies sein soll. In diesem Fall bezeichnet man den Mengenkörper als σ -Körper (oder σ -Algebra) von Mengen (bzw. von Ereignissen).

Sei $(A_i)_{i \in \mathbb{N}}$ die Abkürzung für eine solche Mengenfolge²⁾. Die Definition der unendlichen Vereinigung lautet:

$$\bigcup_{i=1}^{\infty} A_i := \{x \mid \forall i (i \in \mathbb{N} \wedge x \in A_i)\}$$

d.h. $\bigcup_{i=1}^{\infty} A_i$ ist die Menge aller Objekte x , die in mindestens einer der Mengen A_i als Elemente enthalten sind.

Der unendliche Durchschnitt läßt sich wie folgt definieren:

$$\bigcap_{i=1}^{\infty} A_i := \{x \mid \forall i (i \in \mathbb{N} \rightarrow x \in A_i)\}.$$

Wir können somit die Definition des σ -Körpers wie folgt angeben:

D2 Eine Menge Ω von Teilmengen einer gegebenen Menge Q heißt σ -Körper über Q genau dann, wenn gilt:

- (1) $Q \in \Omega$;
- (2) für jedes $E \in \Omega$ ist auch $\bar{E} \in \Omega$;
- (3) für jede Folge $(A_i)_{i \in \mathbb{N}}$ von Mengen aus Ω (d.h. mit $A_i \in \Omega$ für alle i) ist auch $\bigcap_{i=1}^{\infty} A_i \in \Omega$.

Es dürfte klar sein, daß jeder σ -Körper auch ein Körper ist. Die Umkehrung gilt jedoch nicht allgemein³⁾.

1) Es ist leicht zu zeigen, daß auch die leere Menge $\emptyset = \bar{Q}$ in Ω ist. Sind E_i und E_j Elemente von Ω , so ist auch $\bar{E}_i \cup \bar{E}_j = \overline{E_i \cap E_j}$, also der Durchschnitt von E_i und E_j , Element von Ω .

2) $\mathbb{N}$ ist die Menge der natürlichen Zahlen.

3) Wegen $\bigcap_{i=1}^{\infty} \bar{A}_i = \overline{\bigcup_{i=1}^{\infty} A_i}$

ist auch der abzählbar unendliche Durchschnitt der Mengen $A_1, A_2, \dots$ in Ω ; man schreibt dafür auch $\prod_{i=1}^{\infty} A_i$.

Nach der Wahl des Möglichkeitsraumes Q und nach der Wahl eines geeigneten Körpers (bzw. σ -Körpers) von Ereignissen über Q besteht die nächste Aufgabe in der Einführung eines Wahrscheinlichkeitsmaßes, welches für alle Elemente dieses Körpers (bzw. σ -Körpers) definiert ist.

Endlich additive und σ -additive Wahrscheinlichkeitsmaße

Wahrscheinlichkeiten werden, wie früher erwähnt, den Elementen des Körpers bzw. σ -Körpers über Q zugeordnet. Eine Funktion, mittels derer man solche Zuordnung vornimmt, bezeichnen wir mit P . Der Definitionsbereich von P ist also Ω , d.h. $D_I(P) = \Omega$. P ist, wie schon früher festgestellt, eine Mengenfunktion, da die Argumente von P Ereignisse, d.h. Mengen, sind. Die Funktion P heißt Wahrscheinlichkeitsmaß.

Die Einführung von Wahrscheinlichkeiten in der mathematischen Wahrscheinlichkeitstheorie besteht somit aus drei Schritten: Man wählt zuerst einen Möglichkeitsraum Q , dann einen Körper (bzw. σ -Körper) Ω über Q und definiert anschließend eine Mengenfunktion P mit $D_I(P) = \Omega$.

Für die Funktion P gilt, daß sie normiert (d.h. ihre Werte¹⁾ auf das geschlossene Intervall von 0 bis 1 beschränkt sind) und außerdem additiv ist (d.h. für zwei disjunkte bzw. sich wechselseitig ausschließende Ereignisse gilt, daß die Wahrscheinlichkeit, daß entweder das eine oder das andere Ereignis eintritt, gleich der Summe der Einzelwahrscheinlichkeiten dieser beiden Ereignisse ist).

Wir gelangen zur Definition des Wahrscheinlichkeitsraumes, wobei wir die Kolmogoroff-Axiome als Bestandteil dieser Definition anschreiben:

1) P hat als Werte nur reelle Zahlen; daher spricht man von P als einer reellwertigen (Mengen-)Funktion.

D3 W_e ist ein endlich additiver Wahrscheinlichkeitsraum genau dann, wenn es ein Q , ein Ω und ein P gibt, so daß $W_e = \langle Q, \Omega, P \rangle$, wobei folgendes gilt:

- (1) Q ist eine nicht leere Menge;
- (2) Ω ist ein Mengenkörper über Q ;
- (3) P ist eine reelle Mengenfunktion mit $D_I(P) = \Omega$, welche die folgenden Bedingungen erfüllt:
 - A1 Für alle $A \in \Omega$ ist $P(A) \geq 0$;
 - A2 $P(Q) = 1$;
 - A3 Wenn $A, B \in \Omega$ und $A \cap B = \emptyset$, dann $P(A \cup B) = P(A) + P(B)$
 (Forderung der endlichen Additivität, auch spezielles Additionsprinzip genannt).

Verallgemeinerung auf den Fall des σ -additiven Wahrscheinlichkeitsraumes:

D4 W_σ ist ein σ -additiver Wahrscheinlichkeitsraum genau dann, wenn es ein Q , ein Ω und ein P gibt, so daß $W_\sigma = \langle Q, \Omega, P \rangle$, wobei folgendes gilt:

- (1) Q ist eine nicht leere Menge;
- (2) Ω ist ein σ -Körper über Q ;
- (3) P ist eine reelle Mengenfunktion mit $D_I(P) = \Omega$, welche die Bedingungen erfüllt:
 - A1 Für alle $A \in \Omega$ ist $P(A) \geq 0$;
 - A2 $P(Q) = 1$;
 - A3 Für jede unendliche Folge $(A_i)_{i \in \mathbb{N}}$ paarweise disjunkter Mengen A_i , mit $A_i \in \Omega$ für alle $i \in \mathbb{N}$, gilt:

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

(Forderung der σ -Additivität des Wahrscheinlichkeitsmaßes).

Bedingte Wahrscheinlichkeiten und das allgemeine Multiplikationsprinzip

Es sei ein Wahrscheinlichkeitsraum $\langle Q, \Omega, P \rangle$ gegeben. A und B seien Ereignisse, d.h. Elemente von Ω . Ferner sei $P(B) \neq 0$. Die bedingte Wahrscheinlichkeit von B relativ zu A läßt sich wie folgt definieren:

$$P(A/B) := \frac{P(A \cap B)}{P(B)}.$$

$P(A/B)$ ist die Wahrscheinlichkeit von A, wenn bereits die Informationen über das Eintreten des Ereignisses B vorliegen.

Ein Beispiel: Angenommen, in einem Raum befinden sich 6 Frauen und 4 Männer; darunter sind 3 Juristen, also 7 Nichtjuristen. Uns interessiert die Wahrscheinlichkeit für eine beliebige Person, ein Jurist zu sein, unter der Bedingung, daß sie eine Frau ist. Um das feststellen zu können, müssen wir die Anzahl der Juristen, die gleichzeitig Frauen sind, in der Gesamtheit kennen (d.h. den Durchschnitt dieser beiden Ereignisse). Sei diese Anzahl gleich 2. Anhand dieser zusätzlichen Information läßt sich die bedingte Wahrscheinlichkeit ausrechnen. Sie beträgt - da $P(A \cap B) = 2/10$ und $P(A) = 6/10$ - nach der obigen Formel $1/3$.

Aus der Definition gewinnt man das allgemeine Multiplikationsprinzip:

$$(1) P(A \cap B) = P(B) \cdot P(A/B),$$

$$(2) P(A \cap B) = P(A) \cdot P(B/A).$$

Inhaltlich werden die obigen Zusammenhänge folgendermaßen wiedergegeben: Die Wahrscheinlichkeit dafür, daß zwei Ereignisse A und B zusammen eintreten, ist gleich der Wahrscheinlichkeit, daß eines dieser beiden Ereignisse eintritt, multipliziert mit der Wahrscheinlichkeit, daß das andere eintritt unter der Voraussetzung, daß das erste Ereignis vorliegt.

So ist im oben angegebenen Beispiel die Wahrscheinlichkeit dafür, daß eine beliebige Person im Raum sowohl Frau wie auch Jurist sein kann, gleich der Wahrscheinlichkeit, daß sie Jurist (bzw. Frau) ist, multipliziert mit der Wahrscheinlichkeit, daß sie Frau ist unter der Voraussetzung, daß sie Jurist (bzw. Jurist unter der Voraussetzung, daß sie Frau) ist.

Aus diesem Beispiel ersehen wir auch, daß man beide Ereignisse "Frausein" und "Juristsein" (d.h. die beiden Glieder in der bedingten Wahrscheinlichkeit) bei der Berechnung der Wahrscheinlichkeit für das gemeinsame Eintreten beider Ereignisse wegen der Kommutativität von $\cap$ vertauschen kann.

Stochastische Unabhängigkeit von Ereignissen

Angenommen, es gelte $P(A/B) = P(A)$. Dies besagt, daß das Eintreten des Ereignisses B keinen Einfluß auf die Wahrscheinlichkeit des Eintretens von A hat. Man bezeichnet A und B in diesem Fall als (stochastisch) unabhängig. Wir gelangen zur folgenden Definition:

Zwei Ereignisse A und B sind (stochastisch) unabhängig voneinander genau dann, wenn gilt:

$$P(A \cap B) = P(A) \cdot P(B).$$

Als Beispiel für die stochastische Unabhängigkeit können wir die aufeinanderfolgenden Würfe mit einem gleichförmigen Würfel oder mit einer normalen Münze anführen.

Für eine Folge $A_1, A_2, \dots$ von Ereignissen läßt sich der Begriff der Unabhängigkeit so definieren:

Eine Folge $(A_i)_{i \in \mathbb{N}}$ von Ereignissen ist genau dann (stochastisch) unabhängig, wenn jedes A_i unabhängig von jedem Ereignis $\prod_{j \in K} A_j$ ist, d.h. von jedem Durchschnitt von Ereignissen A_k mit $k \neq i$ und beliebig, aber endlich vielen Gliedern. Es gilt:

$$P(A_1 \cap \dots \cap A_n) = P(A_1) \cdot \dots \cdot P(A_n).$$

Das allgemeine Additionsprinzip

Nach A3 gilt, daß $P(A \cup B) = P(A) + P(B)$, falls $A \cap B = \emptyset$ (d.h. der Durchschnitt von A und B leer) ist. Für den Fall, daß das nicht gilt, d.h. $A \cap B \neq \emptyset$, gestattet ein anderes Prinzip die Bestimmung von $P(A \cup B)$.

Dieses Prinzip, genannt das allgemeine Additionsprinzip, lautet:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

(dies ist eine Folgerung aus den Axiomen A1-A3).

Man sieht leicht, daß diese Formel für $P(A \cap B) = 0$ (dies ist der Fall, wenn $A \cap B = \emptyset$) mit dem ursprünglichen speziellen Additionsprinzip identisch ist.

Die Formel der totalen Wahrscheinlichkeit

Diese Formel ist insofern wichtig, als sie uns gestattet, ein für die praktische Anwendung wichtiges Theorem, das sog. Bayes-Theorem (auch Bayes-Formel genannt) herzuleiten.

Um die Formel der totalen Wahrscheinlichkeit herzuleiten, beschränken wir uns auf den Fall des endlich additiven Wahrscheinlichkeitsraumes W_e (der Beweis für W_G verläuft analog).

Es sei eine endliche Folge von n Ereignissen $E_1, \dots, E_n$ gegeben, die eine Zerlegung¹⁾ von Q bilden. A sei ein beliebiges Ereignis. Es gilt:

$$\begin{aligned} A &= A \cap Q = A \cap (E_1 \cup \dots \cup E_n) = (A \cap E_1) \cup (A \cap E_2) \cup \dots \cup (A \cap E_n) = \\ &= \bigcup_{i=1}^n (A \cap E_i). \end{aligned}$$

Daraus folgt:

$$P(A) = P\left(\bigcup_{i=1}^n (A \cap E_i)\right).$$

1) Eine Zerlegung einer gegebenen Menge in n Mengen ist dann gegeben, wenn die gegebene Menge als Vereinigung dieser Mengen darstellbar ist, wobei die letzteren paarweise disjunkt sein sollen, formal: $E_i \cap E_j = \emptyset$ für $i \neq j$ und $E_1 \cup E_2 \cup \dots \cup E_n = Q$. Man verwendet für diesen Zusammenhang auch die Bezeichnung "vollständige Disjunktion".

Nach A3 (weil alle Mengen E_i paarweise disjunkt sind) erhalten wir:

$$P(A) = \sum_{i=1}^n P(A \cap E_i).$$

Ersetzt man jedes Glied $P(A \cap E_i)$ durch seine äquivalente Darstellung mittels des Multiplikationsprinzips, so gelangen wir zu der Formel der totalen Wahrscheinlichkeit:

$$P(A) = \sum_{i=1}^n P(E_i) \cdot P(A/E_i),$$

wobei für alle $i = 1, \dots, n$ gelten soll: $P(E_i) > 0$.

Die Bayes-Formel:

Für eine vollständige Ereignisdisjunktion $E_1, E_2, \dots, E_n$ mit $P(E_i) > 0$ für alle i und jedes Ereignis A mit $P(A) > 0$ gilt:

$$P(E_k/A) = \frac{P(A/E_k) \cdot P(E_k)}{\sum_{i=1}^n P(A/E_i) \cdot P(E_i)} \quad k = 1, 2, \dots, n$$

(Beweis: durch Einsetzen in die vorangegangenen Formeln).

Für $n = 2$ erhalten wir daraus die geläufige Form des Bayes-Theorems:

$$P(E/A) = \frac{P(A/E) \cdot P(E)}{P(A/E) \cdot P(E) + P(A/\bar{E}) \cdot P(\bar{E})}$$

Das Bayes-Theorem ist nicht nur für eine wissenschaftstheoretische Diskussion, sondern auch für die Anwendung in der Rechtspraxis wichtig. Wir werden später (2.Kap.) in die juristische Diskussion dieser Formel einsteigen und außerdem ihre Anwendung auf die Rechtspraxis an Beispielen veranschaulichen (S.125, 212ff).

Zusammenfassung und Ausblick

In diesem Abschnitt haben wir uns darauf beschränkt, ein mathematisches Modell der Wahrscheinlichkeitstheorie möglichst einfach zu schildern. Der Wahrscheinlichkeitsbegriff wurde dabei als ein undefinierter Begriff eingeführt und nur axiomatisch charakterisiert. Durch die Angabe eines Axiomensystems allein können jedoch die Anwendungsprobleme nicht gelöst werden. Dazu bedarf es nämlich einer Interpretation dieses Systems, die die intendierte (vorwis-

senschaftliche) Bedeutung des Ausdrucks "Wahrscheinlichkeit" adäquat wiedergibt. Angesichts dessen, daß es zahlreiche unterschiedliche Formulierungen gibt, in denen der intuitive Wahrscheinlichkeitsbegriff in der einen oder anderen Bedeutung auftritt - Wendungen wie "Glaubwürdigkeit", "Glaubensgrad", "Grad der vernünftigen Erwartung", "Grad der partiellen Wahrheit", "Grad der Nähe zur absoluten Sicherheit", "relative Häufigkeit" und dergleichen - entsteht das Problem, ein genügend umfassendes Explikat anzugeben, das allen diesen Bedeutungen gerecht wird. Im Laufe der Beschäftigung mit diesem Problem in den Grundlagendiskussionen stellte es sich jedoch heraus, daß dieses Vorhaben undurchführbar war, und daß statt einer einzigen Explikation des Wahrscheinlichkeitsbegriffs wir es hier mindestens mit drei verschiedenen zu tun haben. Diese sind: der statistische (oder objektive) Wahrscheinlichkeitsbegriff, mit dessen Explikation die sog. 'objektive' Wahrscheinlichkeitstheorie sich befaßt, der logische Wahrscheinlichkeitsbegriff, der der sog. 'induktiven' Wahrscheinlichkeitstheorie zugrunde liegt (als deren Begründer Carnap gilt) und der subjektive (oder personelle) Wahrscheinlichkeitsbegriff, der der Grundbegriff der sog. 'subjektiven' (oder 'personellen') Wahrscheinlichkeitstheorie ist.

Mit der Darstellung dieser drei Interpretationen des Wahrscheinlichkeitsbegriffs und entsprechend mit drei verschiedenen Ansätzen beim Aufbau einer Wahrscheinlichkeitstheorie werden wir uns im nächsten Abschnitt befassen.

1.2. Verschiedene Interpretationen des Wahrscheinlichkeitsbegriffs

Es gibt verschiedene miteinander konkurrierende Deutungen der mathematischen Wahrscheinlichkeitstheorie. So verstehen Frequentisten (Häufigkeitstheoretiker) unter Wahrscheinlichkeiten relative Häufigkeiten auf lange Sicht. Sie gehen dabei von der Tatsache aus, daß statistische Aussagen als Hypothesen über sog. Zufallsexperimente betrachtet werden, die durch Häufigkeitsauszählungen überprüft werden. Für die Subjektivisten dagegen drückt eine Wahrscheinlichkeit den Grad bzw. die Stärke einer Überzeugung (oder Glaubens) relativ auf verfügbare Informationen aus. Dabei gehen sie von der alltäglichen intuitiven Verwendung des Wahrscheinlichkeitsbegriffs aus: in der Wahrscheinlichkeiten von Einzelereignissen relativ auf die verfügbare Erfahrung (Kenntnisse oder Informationen) beurteilt werden!

Entsprechend sind die beiden Interpretationen mit verschiedenen Aufgaben bei der Explikation des Wahrscheinlichkeitsbegriffs konfrontiert. So besteht die Aufgabe für die Häufigkeitstheoretiker darin, den Zusammenhang zwischen dem Begriff der 'objektiven' Wahrscheinlichkeit und den beobachtbaren relativen Häufigkeiten herzustellen (eine Aufgabe, deren Durchführung bis jetzt nicht gelungen ist). Für die Subjektivisten ist es dagegen wichtig, den vagen Begriff des Überzeugungsgrades (Glaubensgrades) genau zu explizieren und den qualitativen Begriff der Überzeugung zu metrisieren. (Dies ist übrigens den Subjektivisten gelungen: sie konnten eine strenge Begründung der wahrscheinlichkeitstheoretischen Axiome im Rahmen der subjektivistischen Theorie liefern).

Eine Spezialisierung des subjektiven Wahrscheinlichkeitsbegriffs ist der von Carnap entwickelte Begriff der logischen oder induktiven Wahrscheinlichkeit. Neben den üblichen Axiomen muß dieser Wahrscheinlichkeitsbegriff noch eine Reihe zusätzlicher Rationalitätskriterien erfüllen.

Gegenüber früheren Auffassungen, die mit einem Ausschließlichkeitsanspruch auftraten (Reichenbach, v.Mises gingen davon aus, daß der statistische (objektive) Wahrscheinlichkeitsbegriff der 'wahren' Wahrscheinlichkeitsbegriff sei; dagegen vertrat etwa die Finetti die Ansicht, daß man allein mit dem subjektiven Wahrscheinlichkeitsbegriff auskommen könne), ist es heutzutage üblich, beiden Wahrscheinlichkeitsbegriffen eine Existenzberechtigung zuzugestehen, jedoch in verschiedenen Bereichen; So wäre der Anwendungsbereich des statistischen Wahrscheinlichkeitsbegriffs die Physik und die Statistik und der des subjektiven das alltägliche Argumentieren oder allgemeiner: jedes induktive Rasonieren.

1) Gelegentlich verwendet man anstelle der Bezeichnung "subjektiver Wahrscheinlichkeitsbegriff" den Ausdruck "personeller Wahrscheinlichkeitsbegriff", um den normativen Charakter der Theorie hervorzuheben. Die Bezeichnung "subjektiver Wahrscheinlichkeitsbegriff" wird gewöhnlich jedoch nicht nur im engen Sinne eines empirisch-psychologischen Begriffs, sondern in einem umfassenderen verwendet, der sowohl den deskriptiv-empirischen wie auch den normativen Aspekt einschließt. Wir werden daher im folgenden die beiden Begriffe "subjektive" und "personelle" Wahrscheinlichkeit synonym verwenden.

Unser Vorgehen in den nächsten drei Unterabschnitten läßt sich kurz so charakterisieren:

Wir beginnen mit der Schilderung der subjektivistischen Deutung der Wahrscheinlichkeitstheorie. Dabei gehen wir vom alltäglichen intuitiven (vorwissenschaftlichen) Begriff der Wahrscheinlichkeit aus, und zwar in komparativen Wendungen der Art "A ist wahrscheinlicher als B". Diesen Begriff werden wir metrisieren, d.h. ihn in einen quantitativen Wahrscheinlichkeitsbegriff, der den Kolmogoroff-Axiomen genügt, überführen¹⁾. Dadurch wird die Begründung dieser Axiome im subjektivistischen Rahmen geliefert. Danach werden wir auf den Begriff der "rationalen Wetten" eingehen - eine andere, gegenüber der Metrisierung intuitiv plausiblere Möglichkeit, diese Axiome zu begründen. Wir werden außerdem auf den Begriff der "bedingten Wahrscheinlichkeiten" noch einmal (vgl. S.30ff) zurückkommen und sie diesmal subjektivistisch deuten. Die Bildung bedingter Wahrscheinlichkeiten beschreibt den Prozeß des sog. "Lernens aus der Erfahrung", dessen Explikation erstmals im Rahmen der subjektivistischen Theorie gelang. Und schließlich werden wir die sog. Jeffrey-Formel angeben, die gilt, wenn die Informationen, relativ zu welchen die Hypothese beurteilt wird, nicht mit Sicherheit, sondern ihrerseits nur mit Wahrscheinlichkeiten gelten. (Ihre Bedeutung für die juristische Praxis werden wir später, S.212ff erläutern).

Nach der subjektivistischen Deutung werden wir uns der objektivistischen Interpretation zuwenden und uns dabei sowohl die Häufigkeitsinterpretation wie auch die in der neuesten Zeit entstandene Auffassung vom objektiven Wahrscheinlichkeitsbegriff als einem theoretischen Begriff näher anschauen. In theoretischer Hinsicht interessieren uns primär die Schwierigkeiten, die bei der Explikation des statistischen Wahrscheinlichkeitsbegriffs entstehen und die sich auf die objektive Wahrscheinlichkeitstheorie negativ auswirken. Vom praktischen Standpunkt aus (etwa für Physik oder Statistik) ist der Begriff der statistischen Wahrscheinlichkeit nach wie vor von zentraler Bedeutung. Wir werden später (Kap.2) zeigen, daß dieser Begriff auch in der juristischen Praxis (etwa bei der statistischen Auswertung bestimmter Beweismittel) eine zwar begrenzte, jedoch wichtige Rolle spielt.

Aus historischer und systematischer Sicht ist eine Gegenüberstellung von subjektivem und objektivem Wahrscheinlichkeitsbegriff wichtig. Diese wird anschließend durchgeführt.

Der Begriff der logischen Wahrscheinlichkeit ist im Rahmen einer juristischen Wahrscheinlichkeitstheorie wohl kaum von Interesse. Wir werden jedoch darauf vollständigkeitshalber auch eingehen (wir geben dabei die von Carnap aufgestellten Rationalitätskriterien an).

1) Wir werden jedoch dabei keinen strangen Beweis des "Metrisierungstheorems" liefern; denn dieser ist langwierig und schwierig. Der Leser möge sich hier mit der bloßen Zusicherung zufrieden geben, daß eine solche Überführung des qualitativen in den quantitativen Wahrscheinlichkeitsbegriff funktioniert. Einen genauen Beweis findet man z.B. bei F.v.Kutschera, WTI, S.52-58.

1.2.1 Der subjektive Wahrscheinlichkeitsbegriff

Im Alltag werden wir oft mit Äußerungen folgender Art konfrontiert: "Es ist wahrscheinlicher, daß es morgen regnet, als daß es morgen schneit", "Es ist praktisch sicher, daß die meisten Menschen den Atomkrieg nicht überleben werden", "Es ist fast ausgeschlossen, daß der Verdächtige die Tat begangen hat" usw. Damit wollen wir unseren Glauben an das (bzw. unsere Überzeugung vom) Eintreten bestimmter Ereignisse ausdrücken; in der ersten Wendung etwa wollen wir sagen, daß wir an das Eintreten von Regen eher glauben, mit ihm eher rechnen, als damit, daß es morgen schneit. In einer Hinsicht sind die obigen Äußerungen noch unvollkommen: Gewöhnlich beurteilen wir Wahrscheinlichkeiten von Einzelereignissen - auch wenn uns das nicht immer bewußt ist - relativ auf verfügbare Kenntnisse (Informationen). So beurteilen wir etwa die Wahrscheinlichkeit, daß die meisten Menschen den Atomkrieg nicht überleben werden, aufgrund unserer (mehr oder weniger fundierten) Kenntnisse bestimmter physikalischer Zusammenhänge oder vielleicht einfach aufgrund der in der Öffentlichkeit herrschenden Meinungen und Vorstellungen; die Wahrscheinlichkeit, daß es morgen regnen wird, vielleicht aufgrund eines Wetterberichts oder unserer eigenen Beobachtungen von Naturvorgängen, die uns zu einer bestimmten Wahrscheinlichkeitsannahme berechtigen, usf.

In den meisten Fällen dienen solche intuitiven Bewertungen von Ereignissen mit Wahrscheinlichkeiten für praktische Zwecke: Unser Handeln bzw. Entscheiden hängt sehr oft davon ab, wie wir die Wahrscheinlichkeit beurteilen, die wir einem Ereignis zumessen.

An die alltägliche Verwendung des Wahrscheinlichkeitsbegriffs (und zwar in komparativen Vergleichsfeststellungen, der Art "Es ist mindestens ebenso wahrscheinlich, daß es morgen regnet, wie daß es morgen schneit") knüpfen auch die Subjektivisten¹⁾ bei der Erklärung dieses Begriffs an. Für die Subjektivisten drückt eine Wahrscheinlichkeit den Grad oder die Stärke unserer Überzeugung relativ auf unsere Erfahrungen (bzw. unser Wissen oder die verfügb-

1) Der subjektivistische Wahrscheinlichkeitsbegriff wurde vor allem von Bruno de Finetti und Frank P. Ramsey entwickelt; Ramsey, "The Foundations of Mathematics, and Other Logical Essays" (1931); de Finetti, "La Prévision: Ses Lois Logique, Ses Sources Subjectives", annales de l'Institut Henri Poincaré, Bd. 7 (1937), S.1-68.

baren Informationen) aus. Das Hauptanliegen der Subjektivisten besteht darin, den recht vagen Begriff des Glaubensgrades zu präzisieren bzw. ihn zu metrisieren.

Die intuitiven Wahrscheinlichkeitsbewertungen variieren bekanntlich von Subjekt zu Subjekt. Daher bezeichnet man sie in Abgrenzung zum objektiven Wahrscheinlichkeitsbegriff, der eine durch die Natur gegebene Größe bzw. physikalische Eigenschaft charakterisieren soll, als subjektive oder personelle Wahrscheinlichkeiten. Der wissenschaftliche Begriff der subjektiven Wahrscheinlichkeit stellt ein rationales Modell dieses vorwissenschaftlich-intuitiven Wahrscheinlichkeitsbegriffs dar. Es wird hier aber weder der tatsächliche Glaube irgendeiner Person wiedergegeben noch weitere Prinzipien formuliert, die für diesen tatsächlichen Glauben gelten. Was im Rahmen einer subjektivistischen Wahrscheinlichkeitstheorie angestrebt wird, ist die Aufstellung von Normen für die rationale, d.h. konsistente und konsequente Verknüpfung von Glaubensannahmen miteinander. Daher ist es vom normativen Standpunkt aus praktisch irrelevant, ob die tatsächliche Entwicklung von Glaubensannahmen aus vorgegebenen Glaubensannahmen von der rationalen Glaubensbildung abweicht oder nicht. Will man jedoch auch in der Praxis rational vorgehen, so muß man gewissen Rationalitätskriterien (das sind die Axiome des subjektiven Wahrscheinlichkeitsbegriffs) genügen.

Die oben angesprochene Situation ist mit einer solchen in der Logik vergleichbar. Denn diese beschreibt nicht, wie oder was tatsächlich gedacht wird, sondern stellt eine rationale Theorie des folgenreichtigen Schließens dar. Auch ist es praktisch irrelevant, ob man de facto gegen logische Gesetze verstößt oder nicht. Will man jedoch richtig, d.h. logisch denken, so muß man gewissen Normen, die in dieser Theorie formuliert werden, genügen.

Wie oben erwähnt, gehen die Subjektivisten bei der Explikation des Wahrscheinlichkeitsbegriffs von dem intuitiven komparativen Wahrscheinlichkeitsbegriff aus. Um diesen Vorgang systematisch nachzuvollziehen, müssen wir wiederum an einen bestimmten Formalismus anknüpfen. Wir lassen hier die im Abschnitt 1.1 eingeführte Symbolik gelten. Darüber hinaus führen wir aber noch zusätzliche Bezeichnungen und Definitionen ein.

Im folgenden werden der Begriff der komparativen Wahrscheinlichkeit und der Begriff des komparativen Wahrscheinlichkeitsraumes definiert und die Axiome des komparativen Wahrscheinlichkeitsbegriffs angegeben.

Der komparative Wahrscheinlichkeitsbegriff

Bezeichnungen und Definitionen:

$A, B, C, D, E, \dots$ sind Ereignisse (bzw. Sachverhalte oder Propositionen).

" $A \preceq B$ " ist die Abkürzung für die Aussage "Das Ereignis A ist höchstens so wahrscheinlich wie das Ereignis B "; bei $\preceq$ handelt es sich somit um eine zweistellige Relation, die definiert ist für Ereignisse (Sachverhalte oder Propositionen). Solche Ereignisse werden in der subjektivistischen Wahrscheinlichkeitstheorie durch Sätze dargestellt¹⁾.

Wir können definieren:

$A < B := \neg(B \preceq A)$, d.h. A ist weniger wahrscheinlich als B ;

$A \approx B := A \preceq B \wedge B \preceq A$, d.h. A und B sind gleich wahrscheinlich.

Wir wollen nun die Struktur des intuitiven Wahrscheinlichkeitsbegriffs axiomatisch festlegen.

Folgende Bedingungen sollen vom Explikat des Wahrscheinlichkeitsbegriffs stets erfüllt sein²⁾

A1* $A \preceq B \vee B \preceq A$;

A2* $A \preceq B \wedge B \preceq C \rightarrow A \preceq C$;

A3* $\emptyset \preceq A$;

A4* $A \cap C = B \cap C = \emptyset \rightarrow (A \preceq B \leftrightarrow A \cup C \preceq B \cup C)$;

A5* Ist $A_1, A_2, \dots$ eine Folge von Ereignissen, für die gilt

$A_i \subset A_{i+1}$ für alle $i = 1, 2, \dots$, so gilt $\bigcup_{i=1}^{\infty} A_i \preceq B$, wenn für alle i gilt $A_i \preceq B$.

Das Axiom A1* garantiert stets die Vergleichbarkeit von zwei Ereignissen aus Ω ; das Axiom A2* die Transitivität. A1* und A2* zusammen besagen, daß $\preceq$ eine komparative Relation im Sinn einer Quasi-

1) Eine Darstellung von Ereignissen als Sätze bereitet keine zusätzlichen Schwierigkeiten: Denn jede Boolesche Algebra (siehe, S. 8), wie z.B. die Aussagenlogik, ist isomorph zu einem Mengenkörper, d.h. Propositionen einer aussagenlogischen Sprache über Sachverhalte können immer als Mengen eines Mengenkörpers dargestellt werden.

2) Vgl. F.v.Kutschera, WTI, S.50.

reihe¹⁾ ist.

A3* sagt aus, daß die leere Menge (d.h. das logisch unmögliche oder kontradiktorische Ereignis) minimale Wahrscheinlichkeit hat, d.h. praktisch ausgeschlossen ist²⁾.

A4* und A5* beschreiben, wie sich die Relation $\leq$ gegenüber den Operationen der (endlichen oder unendlichen) Vereinigungsbildung verhält³⁾. A4* drückt die Monotonie von $\leq$ bzgl. der Operation $\cup$ aus. A5* besagt, daß der Grenzwert einer aufsteigenden Folge A_i nicht wahrscheinlicher ist als eine obere Schranke für die Wahrscheinlichkeit der Glieder A_i .

Eine Wahrscheinlichkeitsbewertung, die A1* bis A5* genügt, nennt man kohärent bzw. konsistent. Die Kohärenz ist also die Rationalitätsbedingung für das Explikat des komparativen Wahrscheinlichkeitsbegriffs.

A1*-A5* definieren $\leq$ als eine komparative Wahrscheinlichkeit und stellen die notwendigen Bedingungen für die Existenz eines normierten σ -additiven Wahrscheinlichkeitsmaßes über Ω dar.

Die Begriffe des Möglichkeitsraumes und des Ereignisses sollen nun subjektivistisch gedeutet werden.

Es sei Q der Möglichkeitsraum. Q ist die Menge der logisch möglichen Zustände des Systems, auf das sich die Ereignisse beziehen, Q kann z.B. die Menge aller möglichen Zustände der Welt sein, die für jeden Sachverhalt festlegen, ob er gilt oder nicht. Die Zustände bilden eine vollständige Disjunktion⁴⁾. Unter einem Ereignis E können wir dann jene Teilmenge von Ω verstehen, die genau die Zustände enthält, in denen E gilt. Die Relation $\leq$ ist somit auf einer Menge Ω von Teilmengen des Möglichkeitsraumes Q (d.h. auf einem Ereigniskörper Ω über Q) definiert.

1) Von einer Quasiordnung spricht man, wenn die Relation $\leq$ reflexiv (d.h. es gilt $A \leq A$) und transitiv ist.

2) Ebenso wird man fordern, daß Q (d.h. das Ereignis, das irgendein Zustand aus Q realisiert ist) maximale Wahrscheinlichkeit hat. $A \leq Q$ folgt aber aus den angeführten Axiomen.

3) Man kann, falls $A \cap B = \emptyset$, $A+B$ statt $A \cup B$ schreiben (in diesem Fall sieht A4* so aus: $A \leq B \leftrightarrow A+C \leq B+C$), und ebenso $\bigcup_{i=1}^{\infty} A_i$ für $\bigcup_{i=1}^{\infty} A_i$, falls $A_i \cap A_k = \emptyset$, $i, k \geq 1$, $i \neq k$.

4) vgl. S.17, Fn.1), bei der formalen Einführung des Wahrscheinlichkeitsbegriffs bestand Q aus möglichen Resultaten; wir haben dabei einfachheitshalber an die objektivistische Terminologie angeknüpft, in der es üblich ist, von Resultaten (eines Versuchs) zu sprechen. In der subjektivistischen Theorie werden die Elemente von Q subjektivistisch - als Zustände (oder aber auch als Umstände) - charakterisiert.

Wir führen jetzt den Begriff des komparativen Wahrscheinlichkeitsraumes ein.

D5 W_k ist ein komparativer Wahrscheinlichkeitsraum genau dann, wenn es ein Q , ein Ω und eine Relation $\preceq$ gibt, so daß

$W_k = \langle Q, \Omega, \preceq \rangle$, wobei folgendes gilt:

- (1) Q ist eine nicht leere Menge (Möglichkeitsraum);
- (2) Ω ist ein Ereigniskörper (bzw. ein σ -Körper) über Q ;
- (3) $\preceq$ ist eine zweistellige Relation, die definiert ist für Ereignisse aus Ω und für die die Bedingungen A1* bis A5* erfüllt sind.

Die Metrisierung des komparativen Begriffs¹⁾

Die Metrisierung des komparativen Wahrscheinlichkeitsbegriffs besteht darin, daß man zeigt, daß ein Wahrscheinlichkeitsmaß P strikt verträglich ist, d.h. daß gilt:

$$A \preceq B \leftrightarrow P(A) \leq P(B).$$

Es gibt verschiedene Ansätze, den komparativen Wahrscheinlichkeitsbegriff zu metrisieren. Wir werden hierauf nicht im Detail eingehen, da die Beweise in der Regel ziemlich langwierig und schwierig sind, wollen jedoch das Vorgehen kurz charakterisieren.

So benutzt v. Kutschera (dessen Darstellung in WTI wir der obigen Schilderung zugrunde gelegt haben) den Begriff des "ausgezeichneten Wahrscheinlichkeitsraumes", für den er einen Metrisierungssatz formuliert, und gibt dann ein Postulat an, nach dem man jeden komparativen Wahrscheinlichkeitsraum in einen solchen ausgezeichneten Raum einbetten kann, um so zu einer Metrisierbarkeitsbehauptung für alle komparativen Wahrscheinlichkeitsräume zu gelangen.

Savage dagegen hat gezeigt, daß ein endlich additives Wahrscheinlichkeitsmaß P strikt verträglich ist, insofern die komparative Wahrscheinlichkeit "fein und fest" ist.

Zu den notwendigen Bedingungen A1*-A4* gibt Savage eine zusätzliche hinreichende an. Diese lautet:

A6* $A < B$ + es gibt eine endliche Zerlegung $\{C_1, \dots, C_m\}$ von Q mit
 $A \cup C_i < B$ für $i = 1, \dots, m$.

A1*-A4* und A6* garantieren die Existenz und Eindeutigkeit eines normierten endlich additiven Wahrscheinlichkeitsmaßes über Ω . (Savage formuliert die Axiome nur für den endlichen Fall, da er die σ -Additivität ablehnt).

Diesen Ansätzen ist gemeinsam, daß sie ein Metrisierungstheorem angeben, in dem die Existenz eines quantitativen normierten Wahrschein-

1) Es sei hervorgehoben, daß die axiomatische Behandlung der komparativen Wahrscheinlichkeit nicht notwendig gebunden ist an eine subjektive Interpretation der Wahrscheinlichkeit. Es werden jedoch qualitative Wahrscheinlichkeiten plausibel im subjektivistischen Rahmen begründet.

lichkeitsmaßes, das mit dem komparativen Wahrscheinlichkeitsbegriff verträglich ist, behauptet wird. Neben der Existenz eines strikt verträglichen Wahrscheinlichkeitsmaßes wird von Savage auch dessen Eindeutigkeit bewiesen¹⁾.

Es gilt das folgende Metrisierungstheorem:

MT Zu jedem komparativen Wahrscheinlichkeitsraum $\langle Q, \Omega, \preceq \rangle$ gibt es eine reellwertige Funktion P auf Ω mit den Eigenschaften

$$A1 \quad P(A) \geq 0;$$

$$A2 \quad P(Q) = 1;$$

$$A3 \quad P(A \cup B) = P(A) + P(B) \text{ für } A \cap B = \emptyset; \text{ bzw.}$$

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) \text{ für jede unendliche Folge } (A_i)_{i \in \mathbb{N}} \text{ paarweise disjunkter Mengen } A_i, \text{ mit } A_i \in \Omega \text{ für alle } i \in \mathbb{N};$$

so daß P dem komparativen Begriff $\preceq$ entspricht, d.h. es gilt
 $A \preceq B \leftrightarrow P(A) \leq P(B)$ ²⁾

Demgegenüber gilt bei Savage:

MT* Zu jedem komparativen Wahrscheinlichkeitsraum $\langle Q, \Omega, \preceq \rangle$ gibt es genau eine reellwertige Funktion P auf Ω mit den Eigenschaften

$$A1 \quad P(A) \geq 0;$$

$$A2 \quad P(Q) = 1;$$

$$A3 \quad P(A \cup B) = P(A) + P(B) \text{ für } A \cap B = \emptyset;$$

so daß P dem komparativen Begriff $\preceq$ entspricht, d.h. es gilt
 $A \preceq B \leftrightarrow P(A) \leq P(B)$ ³⁾

Wie bereits im Abschnitt 1.1 festgestellt, bezeichnet man die Funktion P auf Ω , die den Eigenschaften A1-A3 genügt, als normiertes Wahrscheinlichkeitsmaß auf Ω .

Jeder komparative Wahrscheinlichkeitsbegriff läßt sich also in Form eines normierten Wahrscheinlichkeitsmaßes metrisieren.

Umgekehrt kann man von einem normierten Wahrscheinlichkeitsmaß auf Ω mittels der Korrespondenzbedingung $A \preceq B \leftrightarrow P(A) \leq P(B)$ zu einem komparativen Wahrscheinlichkeitsbegriff auf Ω übergehen. Man kann nämlich leicht zeigen, daß aus A1 bis A3 folgt, daß ein so definierter Wahrscheinlichkeitsbegriff die Axiome A1* bis A5* erfüllt. Somit lassen sich alle normierten Maße über Ereigniskörpern als metrische Wahrscheinlichkeitsbegriffe auffassen.

1) während bei v. Kutschera das letztere nur in bezug auf einen ausgezeichneten komparativen Wahrscheinlichkeitsraum gilt; vgl. v. Kutschera, S. 53ff.

Einen weiteren Metrisierungsansatz findet man bei R.D. Luce, Sufficient conditions for the existence of a finitely additive probability measure, in: 38 Ann. Math. Statist. (1967), S. 780-786.

2) Vgl. S. 23 Fn. 1) dieser Arbeit.

3) Der Beweis dieses MT ist z.B. bei Fishburn, Utility Theory for Decision Making (1970), S. 192ff zu finden.

Der quantitative Wahrscheinlichkeitsbegriff läßt sich damit intuitiv begründen und als quantitative Form des präzisierten, komparativen intuitiven Wahrscheinlichkeitsbegriffs interpretieren.

Die Axiome A1 bis A3 (die im Abschnitt 1.1 lediglich rein formal als Bestandteil der Definition des (endlich bzw. σ -) additiven Wahrscheinlichkeitsraumes angegeben wurden) stellen also die Axiome des quantitativen Begriffs der subjektiven Wahrscheinlichkeit dar.

Rationale Wetten

Die Axiome des quantitativen Wahrscheinlichkeitsbegriffs lassen sich aber auch auf einem anderen Wege begründen: Man kann nämlich den Wahrscheinlichkeitsgrad, den eine Person einem bestimmten Ereignis beimißt, dadurch ermitteln, daß man Bedingungen untersucht, unter denen sie bereit wäre, auf dieses Ereignis zu wetten. Es läßt sich somit ein Zusammenhang herstellen zwischen der Funktion P , die auf einem Ereigniskörper Ω definiert ist, und der Bereitschaft einer Person, Wetten auf Ereignisse aus Ω mit bestimmten Einsätzen einzugehen. Diesen Zugang zu den Axiomen hat erstmals Ramsey gewählt.

Im Unterschied zur Begründung der Axiome über die Metrisierung eines komparativen Wahrscheinlichkeitsbegriffs, die eine größere allgemeine und systematische Bedeutung hat, ist die Einführung der Wahrscheinlichkeitsaxiome durch die Betrachtung rationaler Wetten intuitiv plausibler. Wir wollen hier darauf kurz eingehen.

Wetten bilden ein einfaches Beispiel dafür, wie unser Handeln von der Wahrscheinlichkeit abhängt, die wir Ereignissen zumessen, von deren Eintreten Erfolg bzw. Mißerfolg einer Handlung abhängt. So besteht anscheinend ein enger Zusammenhang zwischen der Wahrscheinlichkeit, die eine Person X einem Ereignis A zuordnet, und ihrer Bereitschaft, mit einem bestimmten Einsatz auf A zu wetten. Wir wollen den Begriff der "Wette" etwas genauer charakterisieren.

Unter einer Wette von X auf A versteht man einen Vertrag von X mit einer anderen Person Y , ihm einen Betrag s_1 zu zahlen, falls A eintritt, und einen Betrag s_2 von Y zu erhalten, falls A eintritt.

Man bezeichnet $s = s_1 + s_2$ dabei als (Gesamt-) Einsatz der Wette. Dieser Einsatz ist immer positiv. Unter dem Wettquotienten von X bei der Wette auf A versteht man den folgenden Quotienten:

$$q_1(A) = \frac{s_1}{s}.$$

Es gilt $0 \leq q_1 \leq 1$.

Wir identifizieren im folgenden den Geldwert der Beträge s_1 , s_2 und s mit deren subjektivem Wert (oder Nutzen). Wir können somit den zu erwartenden Wert des Gewinnes für X bei einer Wette auf A ausrechnen. Dieser ist $s_2 \cdot P(A) - s_1 \cdot P(\bar{A}) = s \cdot P(A) - s_1 = s \cdot (P(A) - q_1(A))$. Der Erwartungswert des Gewinnes für X ist genau dann größer als 0, wenn $q_1(A) \leq P(A)$. Somit wird X , falls er vernünftig ist, eine Wette auf A nur dann akzeptieren, wenn diese Bedingung erfüllt ist. Unter einer fairen Wette versteht man eine solche, bei deren Eingehen X nicht nur seine eigene Rolle übernimmt, sondern auch seine Rolle mit der seines Partners zu tauschen bereit ist, d.h. wenn er sowohl auf A mit $q_1(A)$ wie auch auf $\bar{A}$ mit $q_2(\bar{A}) = 1 - q_1(A)$ zu wetten bereit ist.

Das ist aber genau dann der Fall, wenn der Erwartungswert des Gewinnes für X und Y gleich ist, d.h. für $q_1(A) = P(A)$. Falls dies erfüllt ist, bezeichnet man einen solchen Wettquotienten q ebenfalls als fair. Der faire Wettquotient kann somit als Maß für dessen subjektive Wahrscheinlichkeit angesehen werden. Auf diese Weise können wir aus den Bedingungen eines vernünftigen Wettverhaltens Aussagen über die Struktur eines subjektiven Wahrscheinlichkeitsbegriffs gewinnen.

Im folgenden wollen wir die Axiome für den quantitativen Wahrscheinlichkeitsbegriff herleiten.

Es sei Ω ein endlicher Ereigniskörper. Für die Ereignisse aus Ω sei ein Wettverhalten festgelegt. Das ist genau dann der Fall, wenn für jedes A ein Wettquotient $q(A)$ festgelegt ist. Eine Wette von X auf A soll die Wette des Partners Y bei der Wette von X auf A sein; d.h. es gilt: $q(\bar{A}) = 1 - q(A)$.

Wir definieren nun, was wir unter einem rationalen (bzw. 'fairen') Wettverhalten verstehen:

Man nennt die auf Ω definierte Funktion q (bzw. das durch sie festgelegte Wettverhalten) rational, wenn es keine (endliche) Menge von Wetten mit durch q bestimmten Wettquotienten gibt, die mit Sicherheit einen Verlust für X ergeben.

Die Forderung der Rationalität des Wettverhaltens ist eine Mindestanforderung, die an einen vernünftigen Wetter gestellt wird.

Es gilt der folgende Satz:

Eine Funktion q ist rational genau dann, wenn sie die Bedingungen erfüllt:

- (1) $0 \leq q(A) \leq 1$;
- (2) $q(Q) = 1$;
- (3) $q(A \cup B) = q(A) + q(B)$ für $A \cap B = \emptyset$.

Dieser Satz besagt, daß die Axiome des metrischen Wahrscheinlichkeitsbegriffs, die wir oben aus den Axiomen des komparativen Wahrscheinlichkeitsbegriffs gewonnen haben, gerade die Mindestanforderung darstellen, die an einen vernünftigen metrischen Begriff der subjektiven Wahrscheinlichkeit zu stellen sind).

(Der Beweis des obigen Satzes ist in v.Kutschera, S.71f zu finden; es sei darauf hingewiesen, daß der erste Teil des Satzes, d.h. der Schluß von der Rationalität von q auf die Bedingungen (1)–(3), zuerst von Ramsey formuliert und von de Finetti bewiesen wurde; den zweiten Teil hat Lehman bewiesen und unabhängig von ihm Kemeny).

Man kann übrigens eine noch stärkere als die oben angeführte Forderung, die der sog. strikten Rationalität, angeben:

Man bezeichnet die auf Ω definierte Funktion q bzw. das durch sie festgelegte Wettverhalten als strikt rational, wenn es keine endliche Menge von Wetten mit Wettquotienten q gibt, die für mindestens ein mögliches Ergebnis einen Verlust und für kein Ergebnis einen Gewinn ergeben.

Ein strikt rationales q ist offenbar auch rational. Es gilt dann der Satz:

Eine rationale Wettfunktion q ist strikt rational genau dann, wenn gilt:

$$q(A) = 1 \rightarrow A = Q.$$

(Beweis, v.Kutschera, S.73).

Bedingte Wahrscheinlichkeiten

Mit dem Begriff der bedingten Wahrscheinlichkeit haben wir uns im Abschnitt 1.1 befaßt, wobei wir uns lediglich für den formalen Aspekt dieses Begriffs interessierten, ohne auf seine inhaltliche Deutung einzugehen. Gerade für die subjektivistische Wahrscheinlichkeitstheorie erweist sich der Begriff der bedingten Wahrscheinlichkeit als sehr wichtig, nicht zuletzt deswegen, weil man in dieser Theorie dadurch den Gedanken des Lernens aus der Erfahrung und darüber hinaus den Begriff der intersubjektiven Übereinstimmung aufgrund gemeinsamer Erfahrungen in befriedigender Weise zu präzisieren imstande ist.

Die Wahrscheinlichkeitsbewertung von Ereignissen, die eine Person X vornimmt, ist von den für diese Ereignisse relevanten Informationen stark abhängig, über welche X verfügt. D.h. eine Wahrscheinlichkeitsbewertung bezieht sich immer auf eine bestimmte Wissenssituation bzw. auf ein bestimmtes Erfahrungsdatum D . So ändert (verkleinert) sich die Wahrscheinlichkeit, die X dem Ereignis beimißt, daß morgen schönes Wetter sein wird, wenn X einen ungünstigen Wetterbericht hört und darüber hinaus im Fernsehen ein entsprechendes Satellitenfoto sieht; oder die Wahrscheinlichkeit, die X dem Ereignis (genauer: dem Sachverhalt) beimißt, daß der Verdächtige die Tat begangen hat, erhöht sich, wenn X erfährt, daß der Verdächtige am Tatort von zwei Zeugen beobachtet wurde, darüber hinaus kein Alibi für die Tatzeit besitzt und außerdem seine Fingerabdrücke an der Tatwaffe festgestellt wurden.

Wir wollen die Abhängigkeit der Wahrscheinlichkeitsbewertung von dem Erfahrungsdatum explizit machen. Wir gehen dabei von einem komparativen Wahrscheinlichkeitsraum $\langle Q, \Omega, \preceq \rangle$ aus, der sich auf ein Datum D bezieht. Uns interessiert die Frage, wie wir diese Wahrscheinlichkeitsbewertung abändern können, wenn wir zusätzlich erfahren, daß das Ereignis E aus Ω eingetroffen ist. $\langle Q, \Omega, \preceq_E \rangle$ ist der aufgrund dieser zusätzlichen Information modifizierte Wahrscheinlichkeitsraum bzw. die durch E bedingte komparative Wahrscheinlichkeit. Es gilt das folgende Prinzip:

$$A7^* \quad A \preceq_E B \leftrightarrow A \cap E \preceq B \cap E.$$

D.h. wenn wir wissen, daß E eingetreten ist, so kommt es bei dem Vergleich der Wahrscheinlichkeiten von A und B einzig und allein auf die in E enthaltenen Teilmengen dieser Ereignisse an, d.h. auf die in A und B enthaltenen Mengen von Zuständen, in denen E gilt.

Wir betrachten die Veränderungen von Wahrscheinlichkeitsbewertungen aufgrund von zusätzlichen Informationen E nur für Ereignisse E mit $E \succ \emptyset$. Denn für $E \approx \emptyset$ folgt für alle Ereignisse A aus Ω $A \approx_E \emptyset$. (Es gilt: $E \approx \emptyset \rightarrow A \cap E \approx \emptyset$ und $E \cap \emptyset \approx \emptyset$, d.h. $A \cap E \approx \emptyset \cap E$, also $A \approx_E \emptyset$). Das Eintreten eines für unmöglich gehaltenen Ereignisses E hat zur Folge, daß die ursprüngliche Wahrscheinlichkeitsbewertung ihrerseits ein unmögliches Ereignis beschreibt.

Es gilt der folgende Satz:

Ist $\langle Q, \Omega, \preceq \rangle$ ein komparativer Wahrscheinlichkeitsraum, so auch $\langle Q, \Omega, \preceq_E \rangle$ mit $E \in \Omega$ und $E \succ \emptyset$.

Man kann leicht zeigen, daß das zweite Tripel den Axiomen $A1^*$ – $A5^*$ genügt, wenn dies auch für das erste Tripel gilt.

Der komparative Begriff $\preceq_E$ läßt sich daher metrisieren. Ist P_E eine zugehörige Metrisierungsfunktion, so schreiben wir für $P_E(A)$ auch $P(A/E)$ und erhalten eine uns schon bekannte Definition der bedingten Wahrscheinlichkeit:

$$P(A/E) := \frac{P(A \cap E)}{P(E)} \quad \text{mit } P(E) > 0.$$

Man kann leicht verifizieren, daß P_E ein normiertes Wahrscheinlichkeitsmaß auf Ω ist, und daß gilt: $P(A/E) \preceq P(B/E) \leftrightarrow P(A \cap E) \preceq P(B \cap E)$, also nach $A7^*$ $A \preceq_E B \leftrightarrow P_E(A) \preceq P_E(B)$. Die so definierte Funktion P_E ist also eine Metrisierungsfunktion für die Bewertung $\preceq_E$.

Die Bildung bedingter Wahrscheinlichkeiten nach $A7^*$ bzw. nach der obigen Definition beschreibt in vereinfachter Form den Prozeß, wie wir unsere Annahmen über die Welt, die sich in einer Wahrscheinlichkeitsbewertung P über Ω ausdrücken, aufgrund von Erfahrungen (Beobachtungen, Kenntnisse und dgl.) abändern. Offensichtlich können wir ohne Erfahrungen keine Wahrscheinlichkeitsbewertung machen, denn es ist anzunehmen, daß unsere Annahmen über die Welt immer auf gewissen Kenntnissen (über die Welt) basieren¹⁾.

1) Eine andere Meinung vertritt Carnap, vgl. S. 50 dieser Arbeit.

Dieser Punkt ist insbesondere für die Diskussion der Apriori-Wahrscheinlichkeitsbewertung eines Ereignisses wichtig. Denn auch diese wird in bezug auf irgendwelche Erfahrungen gemacht. Die Apriori-Wahrscheinlichkeitsbewertung muß außerdem - falls es sich nicht um ein logisch unmögliches Ereignis handelt - positiv sein, weil sie sonst nicht mehr durch die Erfahrung korrigiert werden könnte.

Kontrovers ist dieser Standpunkt in der Rechtspraxis. Diejenigen, die die Wahrscheinlichkeitstheorie als geeignete Methode für die Beurteilung strafrechtlicher Sachverhalte ablehnen, berufen sich gerade darauf, daß die Annahme einer positiven Apriori-Wahrscheinlichkeit den juristischen Grundsatz "Der Angeklagte ist solange unschuldig bis er verurteilt wird" verletze (denn diese müsse 0 sein)¹⁾.

Oft wird auch die Meinung vertreten, daß überhaupt die Annahme irgendwelcher Apriori-Wahrscheinlichkeiten den obigen Grundsatz verletze, das Sprechen in Apriori-Wahrscheinlichkeitsurteilen sei unzulässig, ein Richter müsse daher - wenn überhaupt - seine Überzeugung in Form eines intuitiven augenblicklichen Gesamt-(Wahrscheinlichkeits-)Urteils repräsentieren.

Diese Forderung erscheint - wenn sie nicht performativ gemeint ist - unrealistisch: Auch beim besten Willen kann sich der Richter nicht in einen Zustand versetzen, in dem all seine Erfahrung "schweigt", bis er bei der Urteilsverkündung in einer Art Gesamtschau die Eingebung über das Bestehen bzw. Nichtbestehen des Sachverhalts erfährt.

Realistisch dagegen ist - auch in der Rechtspraxis - die Beschreibung des Prozesses der Überzeugungsbildung als einem Prozeß der Angleichung (Präzisierung oder Modifikation) der anfänglichen Bewertung des Sachverhalts durch den Richter (im Lichte irgendwelcher für den Sachverhalt relevanten Informationen und persönlichen Erfahrung des Richters) an das endgültige Urteil aufgrund der vor Gericht gebrachten Tatsachen.

1) Es ist klar, daß die derart argumentierenden Juristen mit der Apriori-Wahrscheinlichkeit eines juristischen Sachverhalts $P(E) = 0$ etwas anderes meinen als es in der Wahrscheinlichkeitstheorie der Fall ist, denn $P(E) = 0$ besagt - wahrscheinlichkeitstheoretisch interpretiert -, daß in bezug auf unsere bisherigen Erfahrungen das Ereignis E logisch unmöglich ist (bzw. praktisch ausgeschlossen ist). Zu behaupten, daß es praktisch ausgeschlossen ist, daß der Angeklagte schuldig ist, wäre aber auch für die Juristen kaum annehmbar.

Gelegentlich wird der obige juristische Grundsatz auch mit der Annahme der Wahrscheinlichkeit von $1/2$ für die Schuld/Unschuld verbunden. Vgl. J.Kaplan, Decision Theory and the Factfinding Process, in: 20 Stanford Law Review (1968), S.1065-1092; S.1085: "But courts do emphasize the presumption of innocence. The instruction on this presumption is more than a direction to the jury not to consider as evidence against the defendant the fact that he has been indicted and is being tried. It is a direction, at the very least, that the jury not assume at the outset of their consideration of the case that the defendant is more likely guilty than not. Thus the jury is being told Q_0 (die Apriori-Wahrscheinlichkeit - Anm. von mir) is certainly not more than one-half".

Den Prozeß der Veränderung unserer Annahmen durch die Bildung bedingter Wahrscheinlichkeiten aufgrund neuer Erfahrungen können wir so darstellen: Wir wählen aufgrund eines Erfahrungsdatums D_0 , das alle unsere Erfahrungen zu einem Zeitpunkt t_0 enthält, eine Wahrscheinlichkeitsbewertung aus. P sei so gewählt, daß gilt $P(A/D_0) = P(A)$ ¹⁾. Wir bilden zu den Zeitpunkten $t_1, t_2, \dots$ die Bewertungen $P(A/D_1), P(A/D_2), \dots$, wobei $D_1, D_2, \dots$ die an $t_1, t_2, \dots$ zusätzlich zu $D_0, D_1, D_2, \dots$ hinzugekommenen Informationen sind. Ist $P_1(A) = P(A/D_1)$, so ist

$$\begin{aligned} P_1(A/D_2) &= \frac{P_1(A \cap D_2)}{P_1(D_2)} = \frac{P(A \cap D_2/D_1)}{P(D_2/D_1)} = \frac{P(A \cap D_1 \cap D_2)}{P(D_1)} \cdot \frac{P(D_1)}{P(D_1 \cap D_2)} \\ &= P(A/D_1 \cap D_2). \end{aligned}$$

Die sukzessive Bildung der bedingten Wahrscheinlichkeiten besteht also darin, daß ein Ereignis A aus Ω zum Zeitpunkt t die Ausgangswahrscheinlichkeit (Apriori-Wahrscheinlichkeit) von A zum Zeitpunkt t_0 , bedingt durch die im Zeitintervall $t-t_0$ hinzugekommenen Erfahrungen, hat.

Die so beschriebene Bildung bedingter Wahrscheinlichkeiten stellt den sog. Prozeß des Lernens aus der Erfahrung dar. Der Wert von $P(A/D)$ hängt von der Ausgangsbewertung P ab, falls A durch D weder verifiziert noch falsifiziert wird. In bestimmten Fällen verschwindet der Einfluß der Ausgangsbewertung für ein hinreichend großes Erfahrungsdatum. Ansonsten bleibt der Einfluß der sich in P manifestierenden subjektiven Vorurteile bestehen. Aufgrund von Erfahrungen können sich allerdings diese Vorurteile bestätigen oder abschwächen²⁾.

Der Lernprozeß aus der Erfahrung findet sein Ende, wenn man Ereignisse mit $P(A) = 0$ oder $P(A) = 1$ bewertet. Denn diese Bewertungen sind nicht korrigierbar. Es gilt dann $P(A/D) = 0$ (bzw. $P(A/D) = 1$) für alle D .

Man formuliert daher eine Rationalitätsforderung, um solche, durch Erfahrung nicht korrigierbare, dogmatische Wahrscheinlichkeitsbewertungen auszuschließen. Undogmatische Wahrscheinlichkeitsbewertungen nennt man regulär. Es ergibt sich somit die folgende Definition:

Def.: Ist Q eine endliche Zustandsmenge, so heißt eine Wahrscheinlichkeitsbewertung P auf einem Ereigniskörper Ω über Q regulär, wenn gilt:

$$P(A) = 0 \rightarrow A = \emptyset \quad \text{für alle } A \in \Omega$$

1) $P(D_0)$ muß daher 1 sein. Aus $P(D_0) = 1$ folgt nämlich $P(A \cap D_0)/P(D_0) = P(A) = P(A/D_0)$.

2) Zum 'Prozeß des Lernens aus der Erfahrung' vgl. auch S.47f dieser Arbeit.

Man betrachtet also bei einer regulären Bewertung nur das logisch unmögliche Ereignis $\emptyset$ als praktisch ausgeschlossen, und entsprechend das logisch notwendige Ereignis Q als praktisch sicher. Denn ist $P(A) = 1$, so muß $P(\bar{A}) = 0$ sein; für reguläres $\bar{A} = \emptyset$ ist also $A = Q$ (Q ist endlich)¹⁾.

Die Jeffrey-Formel

Bis jetzt haben wir ausschließlich den Fall betrachtet, wo eine zusätzliche Information über ein Ereignis E aus Ω besagte, daß E eingetreten ist. Ein anderer Fall ist nicht minder von Interesse, nämlich, wenn E selbst mit einer gewissen Wahrscheinlichkeit $P^*(E)$ gilt. Ist E seinerseits in bezug auf eine weitere (vorliegende) Information F wahrscheinlich, d.h. gilt: $P^*(E) = P(E/F)$, so können wir die Unsicherheit hinsichtlich E eliminieren, indem wir die Wahrscheinlichkeit der Hypothese H , statt in bezug auf E , relativ zu F bestimmen. Die Formel, die eine Erweiterung der Bayes-Formel darstellt, lautet (wir kürzen sie mit " B_{erw} " ab):

$$P(H/F) = P(H/E \wedge F) \cdot P(E/F) + P(H/\bar{E} \wedge F) \cdot P(\bar{E}/F) \quad (B_{\text{erw}})$$

Gesetzt der Fall, wir sind zwar der Ansicht, daß E vom Grad $P^*(E)$ wahrscheinlich ist, sind jedoch nicht imstande, die Bedingung F anzugeben, relativ zu der E mit einer bestimmten Wahrscheinlichkeit gilt; F kann z.B. von der Art sein, daß nicht einmal unsere sprachlichen Mittel ausreichen, um es auszudrücken. Beispiele dieser Art sind: Der Versuch herauszufinden, ob es sich bei einer flüchtig gesehenen Person um einen Bekannten handelt aufgrund irgendwelcher besonderen Merkmale (z.B. bestimmten Kleidungsstücken), deren Wahrnehmen an der Person allerdings selbst unsicher ist; die Suche nach einem Verbrecher aufgrund eines Phantombildes, d.h. einer Gesichtsdarstellung, die selbst mit Unsicherheit gilt; der Versuch festzustellen, in welcher Sprache zwei nebeneinander stehende ältere Damen sich unterhalten, wobei einzelne Wörter wegen vieler störender Geräusche nicht deutlich wahrgenommen werden können usw. In solchen Fällen ist die obige Formel nicht anwendbar.

R.C. Jeffrey²⁾ hat allerdings eine Formel angegeben, mittels derer man eine quantitative Wahrscheinlichkeitsbewertung P von Ereignissen aus Ω abändern kann, wenn zusätzliche Informationen besagen,

1) vgl. v. Kutschera, S. 66f.

2) R.C. Jeffrey, Logik der Entscheidungen (¹1967), S. 208ff.

daß E seinerseits nur mit einer gewissen Wahrscheinlichkeit gilt, wobei die Bedingung, die E wahrscheinlich macht, nicht (ohne weiteres) angebar ist.

Die Jeffrey-Formel lautet:

$$P_{P^*(E)}(A) = P(A/E) \cdot P^*(E) + P(A/\bar{E}) \cdot P^*(\bar{E})$$

für $0 \neq P(E) \neq 1$.

Statt $P_{P^*(E)}(A)$ ist auch eine andere Schreibweise üblich: $P(A//P^*(E))$. Wir setzen dabei voraus, daß $P(E) \neq 0$ und $P(E) \neq 1$, d.h. wird E vorher als sicher oder unmöglich angesehen, so ist $P(A//P^*(E))$ nicht definiert; es gilt dann:

$$P(E//P^*(E)) = 1 \cdot P^*(E) + 0 \cdot P^*(\bar{E}) = P^*(E),$$

für $P^*(E) = P(E)$:

$$P(A//P^*(E)) = P(A/E)P(E) + P(A/\bar{E}) \cdot P(\bar{E}) = P(A \cap E) + P(A \cap \bar{E}) = P(A),$$

für $P^*(E) = 1$:

$$P(A//P^*(E)) = P(A/E) \cdot 1 + P(A/\bar{E}) \cdot 0 = P(A/E).$$

$P(A//P^*(E))$ ist ein normiertes Wahrscheinlichkeitsmaß über Ω , denn man kann leicht zeigen, daß $P(A//P^*(E))$ für $0 \neq P(E) \neq 1$ die Axiome A1 bis A3 erfüllt.

Die Jeffrey-Formel läßt sich auf Fälle verallgemeinern, wo wir es mit n Informationen zu tun haben. Seien $E_1, \dots, E_n$ solche Informationen, für die gilt $P(E_i) \neq 1$ und $P(E_i) \neq 0$, $i = 1, \dots, n$.

Wir betrachten die Menge aller Konjunktionen der Form $A_1 \wedge A_2 \wedge \dots \wedge A_n$, wobei jedes der A_i entweder mit dem entsprechenden E_i oder dessen Negation identisch ist und wobei die gesamte Konjunktion nicht logisch falsch ist. Es kann bis zu 2^n verschiedene Konjunktionen einer solchen Menge geben.

Nehmen wir an, die voneinander verschiedenen Konjunktionen der Menge $E_1, \dots, E_n$ seien $B_1, B_2, \dots, B_m$.

Dann lautet die Jeffrey-Formel wie folgt:

$$P(A//P^*(E_1, \dots, E_n)) = P(A/B_1) \cdot P^*(B_1) + P(A/B_2) \cdot P^*(B_2) + \dots + P(A/B_m) \cdot P^*(B_m).$$

Beispiel:

Sei $n = 3$. Ferner seien E_1, E_2 und E_3 Informationen, die besagen: X trägt einen schwarzen Hut (E_1), X trägt einen braunen Mantel (E_2) und X trägt einen grauen Schal (E_3). Es gibt insgesamt $2^3 = 8$ Kon-

Junktionen der Form $A_1 \wedge A_2 \wedge \dots \wedge A_n$:

$E_1 \wedge E_2 \wedge E_3$, $\neg E_1 \wedge E_2 \wedge E_3$, $E_1 \wedge \neg E_2 \wedge E_3$, $E_1 \wedge E_2 \wedge \neg E_3$, $\neg E_1 \wedge \neg E_2 \wedge E_3$, $\neg E_1 \wedge E_2 \wedge \neg E_3$,
 $E_1 \wedge \neg E_2 \wedge \neg E_3$, $\neg E_1 \wedge \neg E_2 \wedge \neg E_3$.

Nehmen wir an, wir wollen relativ zu den Informationen E_1, E_2 und E_3 die Hypothese beurteilen, daß es sich um einen Bekannten handelt, wobei gilt:

$$P^*(E_1) = p_1 \quad P^*(E_2) = p_2 \quad \text{und} \quad P^*(E_3) = p_3$$

(wir nehmen an, daß wir X ganz kurz aus dem fahrenden Auto bzw. in der Ferne bei Nacht sehen, so daß wir bzgl. des Vorliegens der Merkmale E_1, E_2 und E_3 unsicher sind).

Ferner sei angenommen, daß $p_1 \neq p_2 \neq p_3$ ist, d.h. der Grad der Sicherheit soll im vorliegenden Fall je nach dem Merkmal verschiedenen empfunden werden.

Die Jeffrey-Formel enthält hier 8 Additionsglieder. Statt sie einzeln auszurechnen, können wir die sieben Konjunktionen: $\neg E_1 \wedge E_2 \wedge E_3$, $E_1 \wedge \neg E_2 \wedge E_3$, ..., $\neg E_1 \wedge \neg E_2 \wedge \neg E_3$ zu einer einzigen Konjunktion $\neg(E_1 \wedge E_2 \wedge E_3)$ zusammenziehen. $\neg(E_1 \wedge E_2 \wedge E_3)$ besagt, daß mindestens eines der Merkmale nicht vorliegt. Wir erhalten:

$$P(A // P^*(E_1 \wedge E_2 \wedge E_3)) = P(A / E_1 \wedge E_2 \wedge E_3) \cdot P^*(E_1 \wedge E_2 \wedge E_3) + P(A / \neg(E_1 \wedge E_2 \wedge E_3)) \cdot P^*(\neg(E_1 \wedge E_2 \wedge E_3)).$$

Seien E_1, E_2 und E_3 voneinander unabhängig, p_1, p_2 und p_3 jeweils 0.9, 0.8 und 0.4 und $P(A / E_1 \wedge E_2 \wedge E_3) = 0.7$, $P(A / \neg(E_1 \wedge E_2 \wedge E_3)) = 0.01$, dann folgt:

$$P(A // P^*(E_1 \wedge E_2 \wedge E_3)) = 0.7 \cdot 0.288 + 0.01 \cdot 0.712 \approx 0.21.$$

1.2.2 Der objektive (bzw. statistische) Wahrscheinlichkeitsbegriff

Während man bei der Deutung von Wahrscheinlichkeiten in der subjektivistischen Theorie an den vorwissenschaftlichen intuitiven Begriff anknüpft (so daß der subjektive Wahrscheinlichkeitsbegriff als sein Explikat ein wohlinterpretierter Begriff ist), bereitet die Interpretation des objektiven Wahrscheinlichkeitsbegriffs gewisse Schwierigkeiten.

Über folgende Punkte ist man sich jedoch einig¹⁾

1. Der statistische Wahrscheinlichkeitsbegriff soll eine objektive Eigenschaft von Dingen oder von Ereignissen (bzw. von Systemen aus solchen) betreffen und soll somit, im Gegensatz zu einem subjektiven Wahrscheinlichkeitsbegriff, nichts mit den Erwartungen eines Subjekts zu tun haben. Wenn wir z.B. dem Ergebnis "Zahl" beim Werfen mit einer Münze die objektive Wahrscheinlichkeit $1/2$ zuordnen, so soll diese Zuordnung die Münze selbst bzw. den physikalischen Vorgang beim Werfen der Münze betreffen und eine Eigenschaft oder Disposition von ihr bzw. ihm darstellen.

2. Während die subjektiven Wahrscheinlichkeiten einzelnen und einmaligen Ereignissen zugeordnet werden, setzt man objektive Wahrscheinlichkeiten in Beziehung zu Ereignistypen. Denn man bezieht sich bei Aussagen über objektive Wahrscheinlichkeiten auf die sog. beobachtbaren relativen Häufigkeiten des Auftretens eines Ereignisses. Dies hat aber nur dann Sinn, wenn man von einem beliebig oft wiederholbaren Prozeß, einem Massenphänomen oder von einem prinzipiell beliebig oft realisierbaren Versuch ausgeht, dessen einzelne Realisierungen voneinander physikalisch unabhängig sind (wie z.B. das Werfen einer Münze, das sich prinzipiell beliebig oft wiederholen läßt, oder ein Massenphänomen wie der Zerfall einer Menge von Radiumatomen). Die möglichen Ergebnisse solcher Prozesse oder Versuche bilden Ereignistypen, denen objektive Wahrscheinlichkeiten zugeordnet werden können.

Es sind somit Massenvorgänge beliebiger Art und nicht singuläre Ereignisse (denn in bezug auf die letzteren kann man nicht von einer relativen Häufigkeit sprechen), die den Gegenstand der Wahrscheinlichkeitsbeurteilung in der objektiven Wahrscheinlichkeitstheorie bilden.

1) vgl. v.Kutschera, S.88ff.

3. Für den Begriff der objektiven Wahrscheinlichkeit sollen die Kolmogoroff-Axiome (vgl. S.15) gelten, d.h. er soll die formalen mathematischen Eigenschaften eines normierten Wahrscheinlichkeitsmaßes p auf einem (σ -) Ereigniskörper (genauer: auf einem Körper von Ereignistypen) haben.

4. Es soll für den Zusammenhang zwischen der objektiven Wahrscheinlichkeit eines Ereignistyps A bei einem Versuch H und der relativen Häufigkeit des Auftretens von A in einer Folge von physikalisch unabhängigen Realisierungen von H gelten:

Für jedes $\varepsilon > 0$ ist es für hinreichend große n praktisch sicher, daß sich die relative Häufigkeit von A in den ersten n Realisierungen von H höchstens um einen Betrag ε von der objektiven Wahrscheinlichkeit $p(A)$ unterscheidet¹⁾.

Wir wollen auf die zwei wichtigsten Interpretationen des objektiven Wahrscheinlichkeitsbegriffs hier kurz eingehen: die Häufigkeitsdefinition und die Interpretation des Wahrscheinlichkeitsbegriffs als einem theoretischen Begriff.

Die Häufigkeitsdefinition

Hier wird der Begriff der objektiven Wahrscheinlichkeit als Häufigkeitsgrenzwert eingeführt. Man unterscheidet dabei zwischen einem endlichen und einem unendlichen Fall. Ist die Gesamtheit der betrachteten Fällen, in denen ein Ereignis A auftreten oder nicht auftreten kann, endlich, so kann man die objektive Wahrscheinlichkeit von A als relative Häufigkeit des Auftretens von A in dieser Gesamtheit bestimmen. Die objektive Wahrscheinlichkeit eines Ereignisses wird dann als Anzahl der günstigen, geteilt durch die Anzahl der möglichen Fälle definiert²⁾. Wenn wir z.B. eine Urne betrachten, die 10 Kugeln enthält, von denen 4 schwarz sind, so ist die Wahrscheinlichkeit $2/5$, eine schwarze Kugel aus der Urne zu ziehen.

Bei unendlichen Gesamtheiten hat es dagegen keinen Sinn, von der Wahrscheinlichkeit eines Ereignistyps als relativer Häufigkeit, d.h. als Anzahl des Auftretens dieses Ereignistyps in der Reihe, geteilt

1) Dieses Prinzip stellt nach der Häufigkeitsdefinition nichts anderes dar als das schwache Gesetz der großen Zahlen, vgl. v.Kutschera, S.79. Auf die Diskussion dieses Prinzips wollen wir hier nicht eingehen; zur weiteren Erläuterung siehe v.Kutschera, S.96f.

2) Vgl. P.S. de Laplace, *Théorie analytique des probabilités*, Paris (1812).

durch die Anzahl der Elemente der Reihe, zu reden¹⁾ Stattdessen kann man, wenn man sich auf eine abzählbare unendliche Folge von Realisierungen $H_1, H_2, \dots$ eines Versuchs (oder Zufallsexperiments) H^* bezieht, davon sprechen, daß die relative Häufigkeit des Auftretens von A in den Ergebnissen $H_1, H_2, \dots$ gegen einen bestimmten Grenzwert konvergiert.

Bei vielen Zufallsexperimenten bzw. Prozessen, die prinzipiell beliebig oft wiederholbar sind, kann man beobachten, daß sich - trotz einer gewissen Unberechenbarkeit in bezug auf die Einzelergebnisse - für eine genügend große Zahl n von Wiederholungen die relativen Häufigkeiten einem festen Wert nähern, von dem sie immer weniger abweichen, je größer man n wählt. Beispiele solcher Zufallsexperimente bzw. Prozesse sind das Werfen mit einer Münze, verschiedene Glücksspiele, physikalische Prozesse, wie etwa der radioaktive Zerfall usw. Das Verhalten solcher Zufallsexperimente bzw. Prozesse auf lange Sicht kann man also dadurch charakterisieren, daß man die relativen Häufigkeiten der möglichen Resultate als gegen einen Grenzwert konvergierend ansieht. Die objektive Wahrscheinlichkeit wird dann mit einem so definierten Grenzwert der relativen Häufigkeit identifiziert.

Der Inhalt der letzten Absätze läßt sich etwas abstrakter, auch so formulieren:

Ausgangspunkt ist ein Zufallsexperiment (Versuch) H , das man beliebig oft auf physikalisch unabhängige Weise wiederholen kann. Weiter sei Ω ein σ -Körper von Ereignistypen über der Menge Q der möglichen Ergebnistypen von H . Die Ereignisse $A_1, A_2, \dots$ des Typs A ($A \in \Omega$) stellen zu den Realisierungen $H_1, H_2, \dots$ von H gehören. Eine Vorschrift bestimmt eine Folge $X = H_1, H_2, \dots$ von Realisierungen von H . Diese Vorschrift besagt, wie die Ergebnisse $x_1, x_2, \dots$ von $H_1, H_2, \dots$ gewonnen werden sollen. Die Vorschrift kann z.B. besagen: "Führe H (in physikalisch unabhängiger Weise) immer wieder durch und notiere alle dabei auftretenden Ergebnisse (in der Reihenfolge ihres Vorkommens)". In diesem Fall nennen wir X eine Grundfolge (von Realisierungen) von H . Entsprechend spricht man von einer zu einer Grundfolge X gehörenden Ergebnisfolge $x = x_1, x_2, \dots$ auch als einer Grundfolge (von Ergebnissen) von H .

Es sei $h_n(A, x)$ die relative Häufigkeit von A in den ersten n Ergebnissen der Folge x . Die Häufigkeitsdefinition lautet dann:

Es sei angenommen, daß für alle Grundfolgen x von H die Grenzwerte $\lim_{n \rightarrow \infty} h_n(A, x)$ existieren und den gemeinsamen Wert r haben.

Dann gilt:

$p(A) := r$.

1) Denn aus mathematischen Gründen hat dieses Verhältnis in der Regel den Wert 1.

Nach dieser bedingten Definition ist die objektive Wahrscheinlichkeit $p(A)$ immer nur unter Bezugnahme auf einen prinzipiell beliebig oft physikalisch unabhängig wiederholbaren Versuch H definiert, wobei die $h_n(A, x)$ für alle Grundfolgen x von H gegen denselben Grenzwert konvergieren müssen.

Der so definierte Begriff der objektiven Wahrscheinlichkeit erfüllt die Axiome A1 bis A3¹⁾.

Die Häufigkeitsdefinition der objektiven Wahrscheinlichkeit ist gegenwärtig sehr umstritten. Sie wird sogar von vielen Wahrscheinlichkeitstheoretikern als inadäquat abgelehnt. Wir führen im folgenden einige der wichtigsten Einwände gegen diese Definition an:

1. Der Grenzwert, gegen den eine Folge relativer Häufigkeiten konvergiert, läßt sich nicht nach Beobachtung endlich vieler Ergebnisse feststellen. (Wie bekannt, ist jede endliche Folge von Beobachtungen mit der Annahme eines beliebigen Grenzwertes verträglich. Es ist daher schwer einzusehen, wie objektive Wahrscheinlichkeitsaussagen empirisch prüfbar sein sollen).

Dieser Einwand betrifft allerdings nicht die Einführung des Wahrscheinlichkeitsbegriffs, sondern seine praktische Verwendbarkeit für empirisch zu testende Hypothesen.

2. Die korrekte Interpretation von Aussagen über Wahrscheinlichkeiten von Einzelereignissen ist aufgrund der Häufigkeitsinterpretation unmöglich.

Manche Objektivisten - wie v.Mises - haben daher Wahrscheinlichkeitsaussagen über Einzelfälle abgelehnt. Reichenbach dagegen versuchte, den Begriff der Wahrscheinlichkeit von Einzelereignissen auf Umwegen in komplizierter Weise in Sätze über relative Häufigkeitsgrenzwerte von Ereignisfolgen umzudeuten. Dies wird von den Subjektivisten jedoch abgelehnt. Für sie bildet nicht der künstliche Begriff des Häufigkeitsgrenzwertes von Ereignisfolgen, sondern die Wahrscheinlichkeitsaussagen über einzelne Ereignisse den Ausgangspunkt für eine korrekte Explikation des Wahrscheinlichkeitsbegriffs. Allerdings können sich die Limestheoretiker damit verteidigen, daß sie einzig und allein den Begriff der statistischen, nicht der personellen Wahrscheinlichkeit einführen wollen (dessen Anwendungsbereich eben die Physik und die Statistik und nicht der des induktiven (intuitiven bzw. alltäglichen) Rasonierens ist).

1) Beweis ist bei v.Kutschera, S. 94f zu finden.

3. Man stellt die Art und Weise, wie innerhalb der Häufigkeitstheorie mit dem Begriff des Grenzwertes operiert wird, in Frage.

Der mathematische Grenzwert setzt voraus, daß jedes Glied der Folge das folgende eindeutig festlegt bzw. daß die Folge nach einer bestimmten mathematischen Gesetzmäßigkeit gebildet ist. Das ist aber gerade bei der durch einen Zufallsmechanismus erzeugten Folge nicht der Fall: jedes Glied in einer solchen Folge ist zufällig, d.h. man kann sein Auftreten nicht berechnen bzw. vorhersagen. Die Frage stellt sich, ob es überhaupt sinnvoll ist, den Grenzwert auf Zufallsfolgen anzuwenden.

4. Die Definition der objektiven Wahrscheinlichkeit als Grenzwert relativer Häufigkeiten ist zirkulär. Die Argumentation ist hier folgende: In die Definition der objektiven Wahrscheinlichkeit geht als Voraussetzung der Unabhängigkeitsbegriff ein (Die Glieder einer unendlichen Folge von Ereignissen, für die die objektive Wahrscheinlichkeit definiert wird, werden als 'voneinander physikalisch unabhängig' angenommen). Dieser Begriff wird aber selbst unter Verwendung des Wahrscheinlichkeitsbegriffs definiert.

Allerdings könnte man diesem Einwand entgehen, indem man verlangt, daß der bei der Definition der objektiven Wahrscheinlichkeit benutzte physikalische Unabhängigkeitsbegriff nicht identisch sein soll mit dem Begriff der stochastischen Unabhängigkeit. Allerdings ist dann auch der Begriff der physikalischen Unabhängigkeit explanationsbedürftig.¹⁾

Der wichtigste Einwand gegen die statistische Definition des Wahrscheinlichkeitsbegriffs ist folgender²⁾

5. Die Verwendung des mathematischen Konvergenzbegriffs ist - wie Stegmüller vorbringt - fehlerhaft und unzulässig. Denn man wechselt den mathematischen Grenzwertbegriff mit dem Gesetz der großen Zahlen ($\hat{=}$ dem wahrscheinlichkeitstheoretischen Grenzwertbegriff).

Der Gedankengang sei hier kurz erläutert:

Nach der Häufigkeitstheorie wird die objektive Wahrscheinlichkeit als ein Grenzwert der relativen Häufigkeiten definiert, wobei der Ausdruck "Grenzwert" im Sinne der klassischen reellen Analysis verstanden wird.

1) Auf die Diskussion dieser Einwände wollen wir hier weiter nicht eingehen. Vgl. Stegmüller, Bd. IV, Teil D, S. 32-41.

2) Vgl. Stegmüller, ibid. S. 37.

Dieser Begriff ist nach Stegmüller durch den maßtheoretischen Konvergenzbegriff zu ersetzen. Anders ausgedrückt: Die strenge Limesdefinition muß durch das starke Gesetz der großen Zahlen ersetzt werden. Das starke Gesetz der großen Zahlen besagt: Wenn wir es mit einer Bernoullischen Versuchsfolge, d.h. mit einer unendlichen Folge von beliebig oft wiederholbaren, voneinander physikalisch unabhängigen Versuchen mit der statistischen Wahrscheinlichkeit θ für das Eintreten eines Ereignisses E zu tun haben, so daß jedes endliche Anfangsstück der Folge dem Gesetz der Binomialverteilung genügt, dann konvergieren die relativen Häufigkeiten $\frac{x}{n}$ des Eintretens von E bei n Versuchen mit der Wahrscheinlichkeit 1 gegen die statistische Wahrscheinlichkeit θ . Die Wendung "konvergiert" muß daher durch "konvergiert mit Wahrscheinlichkeit 1" ersetzt werden. Für die Definition der objektiven Wahrscheinlichkeit muß somit bereits ein Wahrscheinlichkeitsbegriff vorausgesetzt werden.

Dieser Einwand ist nach Stegmüller entscheidend für die Ablehnung der objektivistischen Wahrscheinlichkeitstheorie, denn er zeigt, daß diese praktische Sicherheit mit logischer Notwendigkeit verwechselt¹⁾ Die objektivistische Wahrscheinlichkeitstheorie kann deshalb nicht durch die Häufigkeitskonvergenz definiert werden, weil die relativen Häufigkeiten nicht mit Notwendigkeit gegen den als "objektive Wahrscheinlichkeit" bezeichneten Wert konvergieren, sondern nur "P-fast sicher"²⁾, wie das Gesetz der großen Zahlen zeigt. Jede Verbesserung der Limesdefinition gerät jedoch in Schwierigkeiten, wenn man den gewöhnlichen durch den wahrscheinlichkeitstheoretischen Konvergenzbegriff im Definiens ersetzt. Denn würde das Wahrscheinlichkeitsmaß, auf welches in der Wendung "P-fast sicher" Bezug genommen wird, mit dem zu definierenden Wahrscheinlichkeitsmaß identifiziert, so wäre die Theorie zirkulär. Falls man dagegen einen anderen Wahrscheinlichkeitsbegriff (etwa einer höheren Ordnung) einführen würde, so geriete die Theorie in einen unendlichen Regreß. Man könnte zwar das in "P-fast sicher" erwähnte Wahrscheinlichkeitsmaß in einer anderen, subjektivistischen, Weise deuten, doch wird dies von den Objektivisten abgelehnt.

1) Vgl. Stegmüller, Bd.IV, Teil D, S.37.

2) "P-fast sicher" ist die Bezeichnung für "mit Wahrscheinlichkeit 1", Stegmüller, ibid.

Der objektive Wahrscheinlichkeitsbegriff als theoretischer Begriff ('Propensity-Deutung')

Im Hinblick auf die Schwierigkeiten der Häufigkeitsdefinition der objektiven Wahrscheinlichkeit kann man fragen, ob es andere Möglichkeiten gibt, Klarheit über den Begriff der objektiven Wahrscheinlichkeit zu erlangen. So vertreten viele Autoren die Auffassung, der Begriff der objektiven Wahrscheinlichkeit sei ein sog. theoretischer Begriff, der sich nicht auf andere Begriffe zurückführen lasse (etwa Popper und Hacking, auf die die "propensity"- Interpretation der objektiven Wahrscheinlichkeit zurückgeht; auch Stegmüller vertritt die Auffassung, daß von der Häufigkeitsdefinition zur 'Propensity'-Deutung übergegangen werden muß, will man an dem Begriff der objektiven Wahrscheinlichkeit weiter festhalten¹⁾).

Nach der Propensity-Interpretation soll die objektive Wahrscheinlichkeit eine gewisse Disposition²⁾ von Versuchsanordnungen oder Prozessen angeben oder beschreiben, die sich nicht explizit durch andere Eigenschaften oder Dispositionen definieren läßt. So beschreiben die Wahrscheinlichkeiten der Ergebnisse "Kopf" und "Zahl" eine Disposition der Münze, die sich beim wiederholten Werfen in den relativen Häufigkeiten des Auftretens manifestiert, und die diese bewirken und erklären soll.

Der Zusammenhang zwischen statistischer Wahrscheinlichkeit und relativer Häufigkeit wäre dann nur ein indirekter, d.h. die "statistische Wahrscheinlichkeit" wäre danach kein durch eine Definition charakterisierter Ausdruck, sondern ein theoretischer Term, der nur indirekt mit beobachtbaren Folgen von Ereignissen, an denen relative Häufigkeiten feststellbar sind, in Zusammenhang stünde.

Auch mit diesem Begriff sind Schwierigkeiten verbunden. So läßt sich u.a. der Zusammenhang zwischen der objektiven Wahrscheinlichkeit als theoretischem Begriff und anderen empirischen Begriffen, speziell mit Beobachtungsbegriffen, nicht präzisieren, da keine Gesetzmäßigkeiten sich dafür angeben lassen. So gesehen können Beob-

1) Vgl. Stegmüller, Bd.IV, Teil D, S.39ff.

2) Objektive Wahrscheinlichkeiten als Dispositionseigenschaften wären demnach strukturell ähnlich Eigenschaften wie 'magnetisch', 'elastisch' usw.

bachtungen, nach denen ein Ereignis eine objektive Wahrscheinlichkeit von einem bestimmten Wert hat, nicht durch andere Beobachtungen auf ihre Richtigkeit hin überprüft werden. Es wird daher von einigen Autoren - wie z.B. v. Kutschera - die Auffassung vertreten, daß derartige Aussagen über objektive Wahrscheinlichkeiten - da sie keine ausreichende empirische Interpretation erfahren - auch wenig Sinn haben.

Die Annahme einer Wahrscheinlichkeitsdisposition soll das Auftreten gewisser relativer Häufigkeiten erklären. Nach Meinung von v. Kutschera tut sie das aber keineswegs bzw. nicht besser als die üblichen physikalischen Eigenschaften einer Versuchsanordnung. Überhaupt erkläre sie weder das Ergebnis eines Versuchs, noch die relativen Häufigkeiten, da sie allein durch diese ermittelt werden könne¹⁾.

Aus diesem Grund sei die Annahme einer objektiven Wahrscheinlichkeit als dispositioneller Eigenschaft, die physikalischen Systemen zukommt, nach v. Kutschera ebenso überflüssig wie ihrem Inhalt nach unpräzise. "Damit bleibt aber, trotz aller Einwände, die gegen sie erhoben worden sind, die Häufigkeitsdefinition der bislang befriedigendste Vorschlag zur Interpretation des Begriffs der objektiven Wahrscheinlichkeit"²⁾.

Gerade an diesen letzten Äußerungen sehen wir, daß die Diskussion um die Interpretation des objektiven Wahrscheinlichkeitsbegriffs längst nicht abgeschlossen ist: Vertritt etwa Stegmüller die Auffassung, daß die Häufigkeitsdeutung des Wahrscheinlichkeitsbegriffs fehlerhaft und unzulässig und daher durch die Interpretation als Wahrscheinlichkeitsdisposition (d.h. durch die Deutung des Begriffs der objektiven Wahrscheinlichkeit als einem theoretischen (dispositionalen) Begriff) zu ersetzen ist, so meint z.B. v. Kutschera, daß gerade dies überflüssig ist und die Häufigkeitsinterpretation die bislang befriedigendste Lösung darstellt. Wer Recht hat, ist beim aktuellen Stand der Diskussion schwer zu sagen.

1) Vgl. v. Kutschera, S. 110.

2) v. Kutschera, S. 109.

Subjektivismus gegen Objektivismus

Abschließend wollen wir kurz auf das Verhältnis der subjektiven zur objektiven Wahrscheinlichkeitstheorie eingehen. Die nachfolgende Schilderung ist vorwiegend historisch-systematischer Art¹⁾

De Finetti, der das Konzept der subjektiven Wahrscheinlichkeit erstmals präzisiert hat, war der erste, der versuchte, das Problem des statistischen Schließens auf eine andere (nicht objektivistische) Weise zu lösen. De Finetti vertrat gegenüber der statistischen Wahrscheinlichkeitstheorie eine ziemlich radikale Position. So war für ihn das Sprechen von 'objektiven' Begriffen wie "Versuchsanordnung", "Versuche vom 'selben Typ'", "Zufallsfolgen", "Unabhängigkeit von Ereignissen" und v.a. von "objektiven Wahrscheinlichkeiten" alles andere als sinnvoll. Für ihn gab es nur eine einzige wahre Wahrscheinlichkeit: den Grad, in dem eine Person an etwas glaubt, und nur eine einzige sinnvolle Interpretation des Wahrscheinlichkeitsbegriffs: die subjektivistische Wahrscheinlichkeitstheorie²⁾

Es sind v.a. zwei Motive, die de Finettis Ablehnung der objektivistischen Auffassung erklären. Das erste Motiv ist die Forderung nach der Entscheidbarkeit von Wahrscheinlichkeitsaussagen. Diese Forderung wird von statistischen Hypothesen im objektivistischen Sinn nicht erfüllt. Denn für die Objektivisten ist das Sprechen von unbekannten Wahrscheinlichkeiten der Normalfall. Man geht hier davon aus, daß wir die Wahrscheinlichkeiten gewöhnlich nicht kennen, so daß wir Hypothesen über sie formulieren müssen und diese zu stützen bzw. zu prüfen versuchen. Für die Subjektivisten gilt dagegen: Eine Wahrscheinlichkeitsaussage ist immer eine Aussage darüber, in welchem Grad eine Person X an etwas glaubt. Daher gibt es für sie keine unbekannten Wahrscheinlichkeiten; auch wenn X sie nicht angeben kann, besagt das lediglich, daß X sich des Glaubensgrades nicht voll bewußt ist. Wenn allerdings X eine rationale Person ist, kann der Subjektivist sofort - durch das Verfahren des Anbietens verschiedener Wetten - diesen Grad bewußt machen.

Das zweite Motiv: Die Entdeckung eines Fehlers in der v.Mises-Reichenbachschen Definition der statistischen Wahrscheinlichkeit. (wir sprachen bereits auf S.41f davon): Wenn x eine Folge von Ergebnissen bei Versuchen einer bestimmten Art ist und $h_n(A, x)$ die relative Häufigkeit der Ereignisse A in den ersten n Gliedern dieser Folge darstellt, so definiert z.B. Reichenbach die statistische Wahrscheinlichkeit von A durch: $p(A) := \lim_{n \rightarrow \infty} h_n(A, x)$. Diese Definition beruht aber, wie Stegmüller zeigt³⁾, auf einer Verwechslung praktischer Sicherheit mit logischer Notwendigkeit. Das bedeutet

1) Vgl. Stegmüller, Bd.IV, Teil D, S.220.

2) Stegmüller beschreibt die Haltung de Finettis gegenüber dem Objektivismus folgendermaßen (S.221): "Der Glaube an eine mit physikalischen Systemen verknüpfte objektive Wahrscheinlichkeit ist (nach de Finetti - Anm. von mir) nichts weiter als ein Spezialfall eines metaphysischen Irrglaubens, der auf einer unberechtigten Ontologisierung und Hypostasierung subjektiver Überzeugungsgrade bzw. Zweifelsgrade beruht".

3) Vgl. Stegmüller, Bd.IV, Teil D, S.222 und S.42 dieser Arbeit.

aber, daß in der obigen Formel die strikte Konvergenz durch die "wahrscheinlichkeitstheoretische Konvergenz" ersetzt werden muß.

Genauer sieht das so aus: Es wird nicht mehr die Behauptung aufgestellt, für jede Folge x gäbe es einen Wert $g(A, x)$, so daß die obige Gleichung (" $g(A, x)$ " steht für " $p(A)$ ") gilt, sondern nur, daß die Konvergenz für fast jede Folge x gelte ("fast jede" ist auf ein Wahrscheinlichkeitsmaß P zu relativieren). Formal ausgedrückt: $P(\{x | \lim_{n \rightarrow \infty} h_n(A, x) = g(A, x)\}) = 1$, d.h. für eine beliebige Folge x konvergiert die relative Häufigkeit der A 's in x mit Wahrscheinlichkeit 1 gegen den Grenzwert $g(A, x)$.

Inhaltlich gesprochen bedeutet das, daß die Vergrößerung des (starken) Gesetzes der großen Zahl, welche die Limestheoretiker ihrer Explikation des Begriffs der statistischen Wahrscheinlichkeit zugrunde legen, und welche in der Ersetzung der wahrscheinlichkeitstheoretischen Konvergenz durch die gewöhnliche Konvergenz besteht, wieder rückgängig gemacht wird¹⁾

Es muß also heißen: "konvergiert fast überall" und nicht "konvergiert", was für die objektivistische Theorie als Häufigkeitsinterpretation einen sehr schwachen Punkt darstellt, sogar für sie "tödlich" ist²⁾

Für die subjektivistische Konzeption entstehen mit der Einführung von $g(A, x)$ hingegen keine Schwierigkeiten. Denn hier wird $g(A, x)$ nicht als Wahrscheinlichkeit interpretiert, sondern lediglich mittels des bereits anderswo zur Verfügung stehenden subjektiven Wahrscheinlichkeitsbegriffs P eingeführt. Für die Subjektivisten stellt sie daher das "fiktive Analogon"³⁾ zum Begriff der objektiven statistischen Wahrscheinlichkeit dar, das lediglich dazu dienen soll, im Rahmen der subjektivistischen Theorie Aussagen über objektive Wahrscheinlichkeiten rekonstruieren zu können. Vor diesem Hintergrund ist nun auch die zentrale These der subjektivistischen Theorie zu verstehen: Alle Wahrscheinlichkeitsaussagen sind letztlich zurückführbar auf subjektive Wahrscheinlichkeitsurteile.

Wir haben früher festgestellt, daß man den Wahrscheinlichkeitsgrad, den eine Person einem bestimmten Ereignis beimißt, feststellen kann, indem man die Bedingungen analysiert, unter denen die Person bereit wäre, auf dieses Ereignis zu wetten. (Wir haben davon gesprochen, daß neben der Metrisierung des komparativen Wahrscheinlichkeitsbegriffs die Herstellung eines Zusammenhangs zwischen der Funktion P , die auf einem Ereigniskörper Ω definiert ist, und der Bereitschaft einer Person, Wetten auf Ereignisse aus Ω mit bestimmten Einsätzen anzuschließen, einen weiteren (intuitiv plausibleren) Zugang zur Begründung der Axiome des quantitativen Wahrscheinlichkeitsbegriffs darstellt):

1) Vgl. Stegmüller, Bd. IV, Teil D, S. 222.

2) Stegmüller, *ibid.*, S. 37; vgl. S. 41 dieser Arbeit.

3) Stegmüller, *ibid.* S. 223.

Genauer betrachtet sieht dies so aus: Es sei X eine Person; a sei ein Geldbetrag, der für X von Nutzen ist. Man stellt X vor die folgende (ausschließende) Alternative: Du mußt entweder auf das Ereignis A wetten und erhältst a , falls A eintritt; hingegen nichts, wenn es nicht eintritt. Oder du mußt auf das Ereignis B wetten und erhältst a , wenn B eintritt; hingegen nichts, wenn B nicht eintritt". Wenn X auf A wettet (bzw. A wählt), so wird das als Kriterium dafür angesehen, daß X A für wahrscheinlicher hält als B ¹⁾

De Finetti, auf den dieser 'Kunstgriff' bei der Metrisierung des komparativen Wahrscheinlichkeitsbegriffs zurückgeht, definierte die subjektive Wahrscheinlichkeit oder den subjektiven Glaubensgrad (einer Person X an ein Ereignis A) operational als maximalen Wettquotienten (zu dem X auf A zu wetten bereit wäre). De Finetti hat erstmals gezeigt, daß diese Metrisierung die Kolmogoroff-Axiome (mit Ausnahme der σ -Additivität) erfüllt, d.h. die subjektive Wahrscheinlichkeit erweist sich als ein normiertes Maß über einem Ereigniskörper.

Später wurde nicht mehr mit einzelnen Wetten, sondern mit ganzen Systemen von Wetten gearbeitet (Kemeny, Shimony, Lehman) und ein neues Rationalitätskriterium, genannt Kohärenz, formuliert: Wenn man zugibt, daß es unvernünftig ist, ein System von Wetten anzunehmen, die in jedem Fall mit Sicherheit einen Gesamtverlust ergeben, dann muß man, sofern man die Wahrscheinlichkeit als rationalen Wettquotienten interpretiert, außerdem zugeben, daß die Kolmogoroff-Axiome gelten.

De Finetti gelang es auch, ein Theorem zu beweisen, daß erkenntnistheoretisch sehr wichtig ist. Aufgrund dieses Theorems ergibt sich folgendes: Erstens ist das Sprechen von objektiven Wahrscheinlichkeiten im Rahmen der subjektivistischen Theorie in exakter Weise rekonstruierbar; zweitens kann der Gedanke des Lernens aus der Erfahrung aufgrund dieser Theorie in vernünftiger Weise präzisiert werden; und drittens gestattet diese Theorie, den Gedanken der rationalen intersubjektiven Übereinstimmung aufgrund gemeinsamer Erfahrungen in befriedigender Weise zu behandeln.

Wir wollen hier nur auf eine kurze intuitive Schilderung der beiden zuletzt genannten Ergebnisse eingehen²⁾

Wir gehen davon aus, daß eine Versuchsanordnung H und Versuche von einem bestimmten Typ τ an H gegeben sind. Dabei setzen wir die Vertauschbarkeit von Ereignissen voraus. Die Glieder $E_1, E_2, \dots$ der Folge x von Ereignissen, die durch die Durchführung von Versuchen vom Typ τ an H zustandekommt, heißen vertauschbar, wenn die Wahrscheinlichkeiten aller Konjunktionen von je r Gliedern dieser Folge (r beliebig) gleich sind.

Für jede Person X_i sei eine subjektive Apriori-Wahrscheinlichkeit (Ausgangswahrscheinlichkeit) $w_i^{(0)}$ gegeben. Formal ausgedrückt handelt es sich um ein Wahrscheinlichkeitsmaß, das für den Ereigniskörper über dem Möglichkeitsraum definiert ist, der aus den möglichen Resultaten von Versuchen des Typs τ an H besteht.

1) Es sei hier darauf hingewiesen, daß für die Klärung des Begriffs der Wette es prinzipiell keine Rolle spielt, ob X tatsächlich den Geldbetrag a erhält oder nicht, d.h. ob es sich um realisierbare oder nur hypothetische Wetten ohne Realisierung handelt (wie etwa bei den Wetten auf historische Ereignisse).

2) Vgl. Stegmüller, Bd. IV, Teil D, S. 227.

Da es sich um subjektive Wahrscheinlichkeitsmaße handelt, können die Größen $w_i^{(0)}$ als Wettquotienten bei rationalem (d.h. kohärentem) Wettverhalten interpretiert werden. Diese Anfangswahrscheinlichkeiten können von Person zu Person verschieden sein.

Die Größen $w_i^{(0)}$ liegen fest, bevor mit der Realisierung von Versuchen begonnen wird.

Der Betrag der subjektiven Wahrscheinlichkeit stellt für die Person X_i keine starre Größe dar. X sei rational und setze die Vertauschbarkeit von Gliedern von x voraus. Wir betrachten die relativen Häufigkeiten $h_n(E, x)$ des Auftretens von Ereignissen der Art E in der Folge x bei den ersten n Durchführungen des Versuchs. Die Ausgangswahrscheinlichkeit $w_i^{(0)}$ wird bereits nach der ersten Beobachtung, welche das Resultat $h_n(E, x)$ ergibt, zu dem Betrag $w_i^{(1)}$ abgeändert. Auf diese Weise erhalten wir eine Folge von Aposteriori-Wahrscheinlichkeiten: Ist $w_n^{(n)}$ der rationale Wettquotient, der sich nach Beobachtung von n Realisierungen ergibt, dann konvergiert (unter der Voraussetzung der Vertauschbarkeit) die Folge $w_i^{(0)}, w_i^{(1)}, \dots, w_i^{(m)}, \dots$ mit Wahrscheinlichkeit 1 gegen einen Wert w_i . Auf diese Weise erfolgt das Lernen aus der Erfahrung (die ursprünglichen subjektiven Ausgangswahrscheinlichkeiten werden durch das sich ständig vergrößernde Erfahrungsdatum nach und nach modifiziert und verbessert).

Das in der subjektivistischen Theorie präzisierte Prinzip des Lernens aus der Erfahrung nimmt auf faktische relative Häufigkeitsfeststellungen Bezug.

Wir betrachten jetzt die Person X_j ($i \neq j$). Unter analogen Bedingungen (und nach gewissen weiteren schwachen Voraussetzungen) erhalten wir auch hier die Folge $w_j^{(0)}, w_j^{(1)}, \dots, w_j^{(n)}$, von der gilt, daß sie konvergiert und zwar zu dem selben Grenzwert wie die Folge $w_i^{(k)}$.

Die subjektivistische Theorie kann somit nicht nur den Gedanken präzisieren, daß vernünftige Personen aufgrund von Erfahrungen ihre Vorurteile und vorgefaßten Meinungen schnell berichtigen, so daß der Einklang mit der Erfahrung hergestellt wird, sondern auch erklären, warum Gruppen von vernünftigen Menschen aufgrund gemeinsamer Erfahrungen auch dann zu denselben oder kaum voneinander abweichenden Auffassungen gelangen, wenn ihre subjektiven Meinungen vor der Aneignung von Erfahrungen sehr unterschiedlich waren. (Damit kann auch im Rahmen der subjektivistischen Theorie ein Begriff der Objektivität sinnvoll verwendet werden. Objektivität besteht für die Subjektivisten in der Herstellung intersubjektiver Übereinstimmung aufgrund gemeinsamer Erfahrung).

Kritik an de Finetti

Die radikale Position de Finettis fordert ihrerseits zur Kritik heraus. So werfen manche Opponenten (wie Hacking) der subjektivistischen Theorie u.a. einen programmatischen Charakter vor, der sie für den praktischen Gebrauch wenig geeignet mache, z.B. habe sie bis jetzt kein einziges Beispiel für eine Analyse ihrer Behandlung statistischer Hypothesen und keines einer adäquaten Rekonstruktion probabilistischer Aussagen in den Einzelwissenschaften (wie etwa der Physik) geliefert. Auch de Finettis Forderung nach der Entscheidbarkeit der probabilistischen Hypothesen (das ist eines der beiden

1) Vorausgesetzt natürlich, daß $h_n(E, x)$ konvergiert.

Motive, die de Finetti im wesentlichen zu seiner Ablehnung der objektivistischen Auffassung geführt haben; vgl. S. 45 dieser Arbeit) dürfte mit der Überwindung des generellen Verifikationspositivismus kein Argument gegen die statistische Theorie mehr sein.

Die erste Kritik kann allerdings wiederum rückgängig gemacht werden, wenn man die radikale Haltung de Finettis aufgibt und sich auf den Kompromiß einigt, der der subjektivistischen Theorie einen ganz anderen Anwendungsbereich (exemplarisch den des alltäglichen intuitiven Rasonierens) als der den statistischen Theorie zuweist. Denn auch der statistischen Theorie kann man vorwerfen, daß sie zwar in der Physik oder der Statistik zu plausiblen Ergebnissen führt, in manchen Einzelwissenschaften (in denen das induktive Argumentieren nicht in der Ausdrucksweise von relativen Häufigkeiten gebraucht wird bzw. in ihr rekonstruierbar ist, wie etwa den historischen bzw. Rechtswissenschaften) dagegen keine solche liefert. Die nachfolgende Untersuchung in dieser Arbeit wird uns auch zeigen, daß eine adäquate Rekonstruktion juristischer Aussagen im subjektivistischen Rahmen (und nur in diesem) möglich ist und sogar zu einem besseren Verständnis des induktiv-juristischen Rasonierens führt.

Schwerer wiegt der Einwand (hauptsächlich von Hacking vorgebracht), daß die Änderung von Wettquotienten, in der nach der subjektivistischen Auffassung das Lernen aus der Erfahrung erfolgt, nicht immer dieses adäquat wiedergibt. Hackings Argument läuft darauf hinaus, daß eine Person auch dann z.B. aus der Erfahrung gelernt haben kann, wenn sie zwar ihre Annahmen oder Wettquotienten aufgrund dieser Erfahrung nicht verändert, in diesen aber viel sicherer ist.

Allerdings ist dieses Argument nur dann relevant, wenn es um den Anspruch der subjektivistischen Theorie geht, im Rahmen einer Einzelwissenschaft eine korrekte Lerntheorie aufzustellen. Solche Probleme berühren uns im folgenden aber zum Glück nicht.

Der Anspruch der subjektivistischen Wahrscheinlichkeitstheorie, die (einzig) adäquate Theorie zu sein, ist nach wie vor ungebrochen, wenn es darum geht, probabilistische Aussagen zu rekonstruieren, wobei objektive Verhältnisse nicht feststellbar sind, Informationen nicht in die Form von relativen Häufigkeiten überführbar (bzw. diese dafür überhaupt ungeeignet) sind und wo es primär auf die rationale Überzeugungsbildung ankommt¹⁾.

1) Zur weiteren Kritik an de Finetti bzw. allgemein am Subjektivismus vgl. Stegmüller, Bd. IV, Teil D, S. 232ff.

1.2.3 Der logische Wahrscheinlichkeitsbegriff

Da der logische Wahrscheinlichkeitsbegriff bei der Diskussion der Anwendung der Wahrscheinlichkeitstheorie auf die Rechtspraxis kaum eine Rolle spielt, wollen wir auf ihn nur kurz der Vollständigkeit halber eingehen.

Die systematische Entwicklung eines Begriffs der logischen oder induktiven Wahrscheinlichkeit geht auf R.Carnap zurück¹⁾

Im folgenden werden wir den logischen Wahrscheinlichkeitsbegriff charakterisieren und die von Carnap aufgestellten Rationalitätskriterien angeben, ohne sie einer kritischen Analyse zu unterwerfen.

Die logische Wahrscheinlichkeit wird von Carnap als eine subjektive Wahrscheinlichkeit aufgefaßt, die jedoch zusätzlichen Rationalitätskriterien genügen soll. Nach Carnaps ursprünglicher Ansicht kann durch solche Rationalitätskriterien genau eine Bewertung als die logische Bewertung von Ω aus der Menge aller möglichen subjektiven Wahrscheinlichkeitsbewertungen eines beliebigen Ereigniskörpers Ω ausgezeichnet werden. In seinen späteren Schriften²⁾ ist er sich dessen allerdings nicht mehr so sicher.

Wir wollen angeben, was man unter einer logischen Wahrscheinlichkeitsbewertung von Ω verstehen soll:

Eine logische Wahrscheinlichkeitsbewertung von Ω ist eine subjektive Wahrscheinlichkeitsbewertung von Ω , die eine vollkommen rationale Person vornehmen würde, die frei von Vorurteilen ist und die keine willkürlichen, unbegründeten Annahmen macht³⁾

Carnap geht von bedingten quantitativen Wahrscheinlichkeiten $c(H/E)$ aus. Dabei stellt $c(H/E)$ den Grad des Glaubens einer vollkommen rationalen Person X an das Eintreten des Ereignisses H dar, wenn E das gesamte Erfahrungsdatum (das empirische Gesamtdatum von X) ist. $c(H/T)$ bedeutet dann einen Sachverhalt, der mit logischer Sicherheit gilt, d.h. die Anfangsbewertung von H bzw. die Apriori-Wahrscheinlichkeit von H , noch bevor irgendwelche Erfahrungen gesammelt werden ("T" steht für ein tautologisches Ereignis). $c(H/E)$ sind dabei Funktionen, die nicht für Ereignisse als Mengen, sondern für Sätze H und E erklärt sind, die Ereignisse ausdrücken. Statt eines Ereigniskörpers betrachtet Carnap eine Sprache L , für deren Sätze die Funktionen c definiert sind; $c(H/E) = r$ ist dann ein Satz der Metasprache zu L .⁴⁾

1) R.Carnap, Logical Foundations of Probability (1950), Chicago (1962); The Continuum of Inductive Methods, Chicago (1952).

2) R.Carnap, The aim of inductive logic, in: E.Nagel, P.Suppes und A.Tarski, Logic, Methodology and Philosophy of Science, Stanford (1962), S.303-318.

3) Vgl. *ibid.*, S.303; auch v.Kutschera, S.123.

4) Der Leser, dem die nachfolgende Darstellung zu formal und schwierig erscheint, kann sie ohne Bedenken überspringen. Informelle Ausführungen zum Begriff der logischen (aber auch der subjektiven und objektiven) Wahrscheinlichkeit ist bei E.L.Nell, Wahrscheinlichkeitsurteile in juristischen Entscheidungen (1983), zu finden.

Carnap legt seiner Untersuchung prädikatenlogische Sprachen zugrunde. Es sind Sprachen $\mathcal{L}_N^K$ ($K, N \geq 1$) mit N Gegenstandskonstanten (Abk. "GK") $a_1, \dots, a_N$; auch der Fall abzählbar unendlich vieler GK $a_1, a_2, \dots$ ist zugelassen (hier ist $N = \infty$). K ist hingegen stets endlich. K ist die Anzahl der einstelligen Prädikatenkonstanten (Abk. "PK") $P_1, \dots, P_K$ von $\mathcal{L}_N^K$. Die $a_1, \dots, a_N$ und $P_1, \dots, P_K$ seien die einzigen deskriptiven Konstanten von $\mathcal{L}_N^K$. $\mathcal{L}_N^K$ enthält außerdem als logische Operatoren die logischen Symbole $\neg, \wedge, \vee, \bigwedge_{i=1}^K$ und $\bigvee_{i=1}^K$.

Die Sprachen $\mathcal{L}_N^K$ sind sehr ausdrucksarm, nicht zuletzt deswegen, weil darin mehrstellige PK fehlen.

Carnap nimmt ferner an, daß $\mathcal{L}_N^K$ über einem N -stelligen Objektbereich interpretiert wird, daß es für jedes Objekt eine GK als Namen gibt und daß verschiedene GK verschiedenen Objekten zukommen.

Eine Zustandsbeschreibung in $\mathcal{L}_N^K$ ist ein Satz der Gestalt $\bigwedge_{i=1}^N P_{k_i}(a_i)$, mit k_i aus $1, \dots, K$.

Die Zustandsbeschreibungen sind nicht logisch determiniert; jeder Satz von $\mathcal{L}_N^K$ läßt sich als Disjunktion von Zustandsbeschreibungen charakterisieren. Der Zustandsmenge Q entspricht also die Klasse der Zustandsbeschreibungen, der Klasse Ω (hier: die Potenzmenge von Q) die Menge aller Disjunktionen von Zustandsbeschreibungen, d.h. die Menge aller Sätze von $\mathcal{L}_N^K$.

Das Axiomensystem für die c -Funktionen¹⁾

C1) Ist $\models E=E'$ und $\models H=H'$, so gilt $c(H/E)=c(H'/E')$.

(c soll eine bedingte subjektive Wahrscheinlichkeitsfunktion sein).

Die folgenden Axiome C2 bis C6 lassen sich als Bedingungen für rationale Wetten rechtfertigen.

C2) $0 \leq c(H/E) \leq 1$.

C3) Ist $\models E \supset H$, so gilt $c(H/E)=1$.

C4) Ist $\models E \supset \neg(H \wedge H')$, so gilt $c(H \vee H'/E) = c(H/E) + c(H'/E)$.

C5) Ist $\models E \supset \neg(H_1 \wedge H_k)$ für alle $i, k=1, 2, \dots$ mit $i \neq k$, so gilt $c(\bigvee_{i=1}^{\infty} H_i / E) = \sum_{i=1}^{\infty} c(H_i / E)$.

C6) Gilt $\models E \supset \neg H$ nicht, so ist $c(H \wedge H'/E) = c(H/E) \cdot c(H'/H \wedge E)$.

C7) Ist N endlich und sind H, E Sätze von $\mathcal{L}_N^K$, so gilt
aus $c(H/E)=1$ folgt $\models E \supset H$.

(Dies ist das Regularitätsaxiom, das sich aus der Forderung ergibt, c solle ein System (bedingter) strikt rationaler Wetten bestimmen).

Die Invarianzaxiome lauten:

C8) $c(H/E)$ ist invariant bei Permutation der GK.

C9) $c(H/E)$ ist invariant bei Permutation der PK.

(Diese beiden Axiome erhalten das klassische Indifferenzprinzip bzw. das Prinzip vom unzureichenden Grund).

C10) $c(H/E)$ ist unabhängig von der Zahl N der GK der zugrundeliegenden Sprache $\mathcal{L}_N^K$. (Dieses Axiom legt die Unabhängigkeit des Wertes $c(H/E)$ von der Sprache, in der die durch H und E dargestellten Ereignisse ausgedrückt werden, fest).

Die nächsten beiden Axiome besagen, daß die logische Wahrscheinlichkeit so gewählt werden soll, daß wir aus der Erfahrung lernen können, und zeigen, welche Komponenten der Erfahrung für die Wahrscheinlichkeit eines Satzes relevant sind. Diese sind das Axiom der Relevanz von Einzelfällen und das Axiom der Voraussageirrelevanz.

1) Dieses Axiomensystem ist im wesentlichen in "Probability and Induction" angegeben; vgl. v. Kutschera, S.126-130.

C11) Es sei E ein Satz der Gestalt $\bigwedge_{i=1}^n P_{k_r}(a_{i_r})$, in dem die beiden verschiedenen GK a und b nicht vorkommen.

Für $n=0$ sei E eine Tautologie T. Dann gilt:

(a) $c(P_k(b)/E \wedge P_k(a)) \geq c(P_k(b)/E)$ und

(b) $c(P_k(b)/E \wedge P_k(a)) \neq c(P_k(b)/E)$.

(Dieses Axiom besagt, daß die Wahrscheinlichkeit von $P_k(b)$ mit positiven Instanzen von P_k zunimmt).

C12) E habe die Gestalt $\bigwedge_{i=1}^n P_{k_r}(a_{i_r})$ und enthalte die GK a nicht.

(a) Es sei $\kappa > 2$ und E' gehe aus E hervor durch Ersetzung eines Vorkommnisses der P_k mit $l \neq k$ durch P_j mit $j \neq l$ und $j \neq k$. Dann gilt:

$c(P_k(a)/E) = c(P_k(a)/E')$.

(b) Für $\kappa=2$ ist $(P_k(a)/E)$ für festes n eine lineare Funktion der Anzahl n_k der Vorkommnisse von P_k in E.

(C12 besagt, daß die Wahrscheinlichkeit von $P_k(a)$ nur von der Zahl der positiven und negativen Instanzen von P_k in E abhängt).

Es seien hier kurz noch einige Einwände gegen die Theorie der logischen Wahrscheinlichkeit angeführt:

1. Die Sprachen, für die man die c-Funktionen bisher definiert hat, sind sehr ausdrucksarm, so daß viele wichtige Probleme darin nicht formuliert werden können.

2. Es ist bis jetzt nicht gelungen, die c-Funktionen durch einen bestimmten Parameter λ eindeutig festzulegen.

3. Die Axiome, mit denen die Theorie der logischen Wahrscheinlichkeit über die der subjektiven Wahrscheinlichkeit hinausgeht, sind fragwürdig¹⁾.

4. Der schwerwiegendste Einwand gegen Carnaps Theorie ergibt sich (nach der Auffassung von v.Kutschera) aus der Paradoxie der Induktion, dem sog. 'Goodman-Paradox'. Aufgrund dieser Paradoxie erweisen sich die Carnapschen Axiome als inadäquat.

(Auf die Diskussion dieser Einwände kann im Rahmen dieser Arbeit nicht näher eingegangen werden²⁾).

1) Vgl. v.Kutschera, S.138.

2) Eine ausführliche Diskussion der Einwände findet der Leser bei v.Kutschera, S.137ff. Zu Carnaps Theorie siehe R.Carnap und W.Stegmüller, Induktive Logik und Wahrscheinlichkeit (1959), W.Stegmüller, Carnap II: Normative Theorie des induktiven Rasonierens, in: Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie, Bd.IV, Teil C.

Zusammenfassung und Ausblick

Aus der obigen Darstellung konnten wir ersehen, daß die subjektivistische Wahrscheinlichkeitstheorie gegenüber der objektivistischen und der logischen gewisse unbestreitbare Vorzüge besitzt, die ihr Akzeptieren leichter machen.

Nicht nur kann man innerhalb dieser Theorie den Begriff der (subjektiven) Wahrscheinlichkeit einwandfrei definieren, sondern auch über objektive (statistische) Wahrscheinlichkeiten sprechen. Außerdem lassen sich innerhalb dieser Theorie zwei wichtige Gedanken: derjenige des Lernens aus der Erfahrung¹⁾ und derjenige der rationalen intersubjektiven Übereinstimmung aufgrund gemeinsamer Erfahrungen, präzisieren.

Gegenüber der logischen Wahrscheinlichkeitstheorie kommt sie außerdem mit viel weniger Rationalitätskriterien aus; darüberhinaus vermeidet sie alle Schwierigkeiten, die sich mit der Auswahl einer bestimmten Wahrscheinlichkeitsfunktion aus einer unendlichen Menge solcher Funktionen ergeben.

Angesichts dessen, daß der statistische Wahrscheinlichkeitsbegriff (in der objektivistischen Auffassung) nach wie vor eine zentrale Rolle in der Statistik und der Physik spielt, wäre die völlige Ablehnung der objektivistischen Theorie allerdings voreilig.

Demgegenüber scheint der Standpunkt²⁾, wonach wir es mit zwei verschiedenen Anwendungsbereichen der Wahrscheinlichkeitstheorie (zum einen, wie oben erwähnt, die Statistik und die Physik und zum anderen das induktive (alltägliche, intuitive) Rasonieren) und dementsprechend mit zwei verschiedenen Wahrscheinlichkeitsbegriffen, d.h. einem objektiven und einem subjektiven, zu tun haben, vernünftig zu sein. Dieser Standpunkt wird auch in dieser Arbeit übernommen.

Zum Bereich des induktiven (alltäglichen, intuitiven) Rasonierens gehört auch das Argumentieren in der Rechtspraxis. Es wird hier die Auffassung vertreten, daß man dieses Argumentieren in korrekter Weise als dasjenige in Wahrscheinlichkeitsurteilen rekonstruieren und den juristischen Begriff der Wahrscheinlichkeit mit dem subjektiven Wahrscheinlichkeitsbegriff identifizieren kann. Mit der Begründung dieser Auffassung werden wir uns im nächsten Paragraphen befassen.

1) wobei dieses Ergebnis, wie auf S.49 erwähnt, nicht uneingeschränkt akzeptiert wird.

2) Vgl. Carnap-Stegmüller, Induktive Logik und Wahrscheinlichkeit (1959).

§2. Subjektive Wahrscheinlichkeitstheorie und Rechtspraxis

In diesem Paragraphen wird gezeigt, daß

- das Argumentieren bzw. Schlüsseziehen in der Strafrechtspraxis zum Bereich des induktiven Rasonierens gehört (Abschnitt 2.1);
- das Denken in Wahrheitskategorien für die juristische Praxis nicht adäquat ist (2.2);
- dem induktiven Rasonieren in der Rechtspraxis ein Wahrscheinlichkeitsbegriff zugrunde liegt, der eindeutig als subjektiver Wahrscheinlichkeitsbegriff zu identifizieren ist (2.3).

Ferner werden wir auf die Gründe für das Bestreben eingehen, den Wahrscheinlichkeitsbegriff - wenn davon überhaupt in der juristischen Literatur die Rede ist - als den Begriff der objektiven oder statistischen Wahrscheinlichkeit zu deuten (2.4).

Anschließend werden wir uns mit der juristischen Forderung nach der "völligen Überzeugung des Richters vom Bestehen der Wahrheit" auseinandersetzen und dabei auf das "Rechtsgefühl des Richters" als Beurteilungskriterium wie auch auf die Problematik der Nicht-Berücksichtigung des Gesamtdatums - als einer Konsequenz der obigen Forderung - näher eingehen (2.5).

Außerdem werden wir auf die Auffassung von L.J.Cohen, nach der der juristische Wahrscheinlichkeitsbegriff anders strukturiert sein soll als derjenige, dessen Präzisierung der mathematische Begriff der Wahrscheinlichkeit ist, hinweisen (2.6). (Wir beschränken uns auf einen Hinweis, weil der von Cohen verwendete Formalismus zu aufwendig ist, als daß wir im Rahmen dieser Arbeit auf seine Theorie ausführlicher eingehen könnten. Mit Cohen werden wir uns noch einmal später (Kap.2, §3) befassen: Wir werden dabei die von ihm gegen die Wahrscheinlichkeitstheorie aufgestellten Paradoxien schildern und ihre Lösungen vorstellen).

Und schließlich wird auch die Möglichkeit, juristischen Entscheidungen statt eines Wahrscheinlichkeitsbegriffs einen solchen des "Tatsachengewichts" (Keynes: "weight of evidence") zugrunde zu legen, untersucht und zurückgewiesen (2.7).

In dem vorangegangenen Paragraphen haben wir uns sowohl die Struktur wie auch Definitionen und Anwendungsmöglichkeiten verschiedener Wahrscheinlichkeitsbegriffe angesehen. Vor allem ging es uns dabei um die (formale und inhaltliche) Aufhellung der Begriffe "subjektive" und "objektive" Wahrscheinlichkeit. So haben wir gesehen, daß die beiden Begriffe dieselbe formale Struktur haben, jedoch eine unterschiedliche Interpretation und Anwendung erfahren: So stellt der subjektive Wahrscheinlichkeitsbegriff eine Präzisierung des vorwissenschaftlichen, intuitiven Begriffs der Wahrscheinlichkeit dar, den wir im Alltag verwenden; daher ist sein Anwendungsbereich jegliches Gebiet induktiven (intuitiven) Rasonierens. Der objektive (bzw. statistische) Wahrscheinlichkeitsbegriff hingegen betrifft eine Eigenschaft oder Disposition von Dingen selbst, die durch Auszählung bzw. Überprüfung von relativen Häufigkeiten bei beliebig oft wiederholbaren oder reproduzierbaren Vorgängen feststellbar ist¹⁾. Der Anwendungsbereich des objektiven Wahrscheinlichkeitsbegriffs ist primär der der Physik und der Statistik, da wir es gerade hier oft mit der Beurteilung solcher Massenvorgänge zu tun haben.

1) Auf die Schwierigkeit, den objektiven Wahrscheinlichkeitsbegriff widerspruchsfrei zu definieren und einen Zusammenhang zum Begriff der relativen Häufigkeit herzustellen, sind wir im Abschnitt 1.2.2 näher eingegangen.

Außerdem haben wir festgestellt, daß die Wahrscheinlichkeitstheoretische Beurteilung von Einzelereignissen der Art "Morgen wird es regnen", "Den Atomkrieg überlebt niemand", "Die Amazonen wurden von Griechen vernichtet" usw. nur im Rahmen der subjektiven Wahrscheinlichkeitstheorie möglich ist; d.h. will man korrekte probabilistische Aussagen über vergangene, zukünftige oder allgemeine Einzelereignisse aufgrund der zur Verfügung stehenden Erfahrungen (in Form von Informationen und sonstigen Kenntnissen) machen, so muß man zum Instrumentarium der subjektivistischen Theorie greifen.

Nicht nur im Alltag oder in der historischen Forschung haben wir es mit der Beurteilung von unsicheren Ereignissen zu tun, sondern auch in der Strafrechtspraxis¹⁾ Dies eingehend zu untersuchen, ist unsere gegenwärtige Aufgabe.

2.1. Wir beginnen damit, daß wir uns Gedanken darüber machen, was den Gegenstand der strafrechtlichen Beurteilung im Prozeß bildet.

Auch einem Nicht-Juristen wird klar sein, daß es im Strafprozeß um die Beantwortung der Frage geht, ob ein bestimmter Tatbestand erfüllt oder nicht erfüllt ist, oder konkreter: um den Nachweis, daß der Angeklagte die ihm vorgeworfene Straftat begangen hat bzw. daß er sie nicht begangen hat. Ist diese Frage beantwortet, so wird eine entsprechende Entscheidung gewählt: eine durch die Gesetzgebung für diesen Tatbestand festgestellte Strafe im ersten Fall; Freispruch im zweiten Fall.

Bei dem juristischen Nachweis eines bestimmten Ereignisses (oder Sachverhalts) handelt es sich selbstverständlich um keinen logischen Beweis: Denn die Beziehung zwischen den Prämissen und der Schlußfolgerung ist nicht deduktiv; es wird hier das Ereignis bzw. der Sachverhalt beurteilt aufgrund von etwas - Beweismitteln, Zeugenaussagen, sonstigen relevanten Tatsachen, kurz: aufgrund einer empirischen Informationsbasis -, was nicht notwendigerweise das Bestehen des Sachverhalts impliziert, sondern diesen lediglich glaubwürdiger oder weniger glaubwürdig macht²⁾

1) Die Unsicherheit ist nicht allein für die Strafrechtspraxis charakteristisch, sondern gilt allgemein für die Rechtspraxis.

2) Die normative Begründung (d.h. die Begriffsanalyse) bei der Bestimmung des Tatbestands ist zwar wichtig, wird aber in dieser Darstellung nicht angegangen. Wir befassen uns ausschließlich mit der empirischen Begründung einer juristischen Sachverhaltshypothese der Art "A hat die Straftat B begangen" aufgrund der im Prozeß zur Verfügung stehenden Beweismittel.

Derartige Schlüsse, die nicht mit logischer Notwendigkeit die Schlußfolgerung implizieren, heißen induktive Schlüsse und die Bildung und Begründung solcher Schlüsse gehört in den Bereich des induktiven Argumentierens.

Juristische Schlüsse sind somit induktive Schlüsse. Der Vergleich einer beliebigen juristischen Sachverhaltshypothese der Art "Der Angeklagte hat die Mordtat begangen", "Der Angeklagte hat den Diebstahl begangen" usw. mit einer historischen Hypothese "Griechen haben die Amazonen ausgerottet", einer biologischen "Die Veränderung des Klimas war die primäre Ursache für das Aussterben der Dinosaurier" oder einer allgemeinen "Den Atomkrieg überlebt niemand" (bei den letzten drei handelt es sich zweifellos um typische induktive Hypothesen) macht es deutlicher.

Denn welcher prinzipieller Unterschied besteht zwischen der Beurteilung von Äußerungen der ersten und einer solchen der zweiten Art?¹⁾ Anscheinend keiner: sowohl im Fall einer juristischen Hypothese wie auch in dem einer historischen, biologischen oder allgemein-wissenschaftlichen Hypothese sind es bestimmte empirische Befunde (sog. "Beweismittel" im ersten, Indizien in Form von Sagen und Überlieferungen im zweiten, Fossilienfunde im dritten, Forschungsergebnisse und Tests im vierten Fall) und sonstiges Wissen (in dessen Lichte man die empirischen Befunde interpretiert), welche die Hypothese stützen oder unsicherer machen (bzw. ihre Glaubwürdigkeit vergrößern oder verringern). In keinem der vier Fälle können wir zum Zeitpunkt der Beurteilung aber die Wahrheit feststellen bzw. den Sachverhalt verifizieren. Darüber hinaus handelt es sich in den drei ersten Fällen um prinzipiell weder verifizierbare noch falsifizierbare Hypothesen, im vierten Fall allerdings um eine prinzipiell falsifizierbare Hypothese²⁾.

Gewöhnlich geht man davon aus, daß alle induktiven Schlüsse Wahr-

1) Der Unterschied besteht lediglich zwischen den einzelnen Hypothesen: So beschreiben die ersten vier vergangene Ereignisse, die fünfte hingegen ein allgemeines, quasi-zukünftiges Ereignis.

2) Im juristischen Fall kann in gewissen Situationen eine unmittelbare Zeugenbeobachtung der Straftat vorliegen. Aber auch dann kann von der Wahrheit der juristischen Hypothese kaum die Rede sein. Denn der den Sachverhalt beurteilende Richter bezieht sich bei seiner Beurteilung nur indirekt auf eine solche Beobachtung. Juristische Hypothesen bleiben somit (vom Standpunkt des Richters aus) stets unverifizierbar.

scheinlichkeitsschlüsse sind¹⁾. Sie müssen also als solche rekonstruiert werden, will man rational (d.h. wissenschaftlich) bei der Beurteilung einer Sachverhaltshypothese vorgehen.

Bei den obigen Hypothesen handelt es sich gerade um derartige Wahrscheinlichkeitsurteile bzw. probabilistische Äußerungen, denn es hat einen Sinn, zu fragen, ob es aufgrund des gegenwärtigen Kenntnisstandes (sehr) wahrscheinlich ist, daß den Atomkrieg niemand überlebt; ebenso: ob es aufgrund der vor Gericht gebrachten Beweismittel in einem Prozeß (sehr) wahrscheinlich ist, daß der Angeklagte der Mörder ist, oder ob das eher 'unwahrscheinlich' ist usw.

2.2. Eine andere Frage ist, ob man - obwohl die Wahrheit des Sachverhalts praktisch nicht feststellbar ist - so tun kann, als ob das der Fall wäre²⁾

Angesichts dessen, daß dies das übliche Vorgehen in der Rechtspraxis ist, wobei die Urteilssprüche in Behauptungen der Form "man sei überzeugt" (oder "man sei sicher"), daß "der Sachverhalt wahr ist" oder zumindest "mit an Sicherheit grenzender Wahrscheinlich-

1) Zumindest ist dies der Standpunkt sowohl der personalistischen Schule wie auch Carnaps. Unter dem induktiven Rasonieren wird dabei ein probabilistisches Rasonieren verstanden. Allerdings ist dann der Gegensatz "deduktiv-induktiv" offensichtlich nicht mehr erschöpfend. Denn es sind Schlüsse denkbar, die nach der obigen Einteilung weder logisch noch induktiv-probabilistisch wären. Zu solchen gehören z.B. die aus der Statistik bekannten Likelihood-Schlüsse, die obachon mit Hilfe des Wahrscheinlichkeitsbegriffs definierbar, keine Wahrscheinlichkeitsschlüsse sind. Mehr zum Begriff der Likelihood siehe bei J.Hacking, Logic of Statistical Inference (1965); A.F.Edwards, Likelihood. An Account of the Statistical Concept of Likelihood and its Application to Scientific Inference (1972); W.Stegmüller, Probleme und Resultate etc. Bd.IV, Teil D.

2) Man behauptet nichts Neues, wenn man sagt, daß das Denken in Wahrscheinlichkeits-, anstatt in Wahrheitskategorien den Juristen weniger zusagen dürfte. Vgl. dazu Nell, S.15: "Der Begriff der Wahrscheinlichkeit, der "Ungewißheit als notwendigen Bestandteil impliziert, muß dem Juristen, der sich vorwiegend mit der Eliminierung von Ungewißheit beschäftigt, von vornherein Mißbehagen bereiten" (Zitat von Kühne, NJW 1979, 627 (618) - Anm. in Fn.). In der Tat erscheint das Denken in Wahrscheinlichkeiten auf den ersten Blick als eine Art Fremdkörper in der juristischen Denk- und Entscheidungsweise. Wo etwas nur wahrscheinlich ist, ist immer auch das Unwahrscheinliche möglich. Wer eine Entscheidung auf eine Annahme stützt, die nur wahrscheinlich ist, macht sich das Risiko bewußt, daß sich seine Entscheidung später als "falsch" erweist. Die Minimierung dieses Risikos ist ein legitimes Anliegen und die Pflicht eines jeden, der verantwortlich Entscheidungen zu treffen hat. Ob man das Risiko allerdings dadurch minimieren kann, daß man auf die Berücksichtigung von Wahrscheinlichkeitsurteilen verzichtet und mit Hilfe von Beweisregeln aus Ungewißheit eine Quasi-Gewißheit macht, ist fraglich".

keit¹⁾ festgestellt ist" gekleidet werden, kann man diese Frage bejahen.

Zahlreiche Äußerungen aus der Rechtsprechung belegen, daß das Ziel des Urteilspruchs in der Herstellung der 'Fiktion der Wahrheit' gesehen wird; vgl. den am meisten zitierten Beleg dafür:

"Vermöge der Beschränkung der Mittel menschlicher Erkenntnis kann niemand (selbst im Falle eigener unmittelbarer Anschauung eines Vorgangs) zu einem absolut sicheren Wissen von der Existenz eines Tatbestandes gelangen... Deswegen gilt im praktischen Leben der hohe Grad von Wahrscheinlichkeit, welcher bei möglichst erschöpfender und gewissenhafter Anwendung der vorhandenen Mittel der Erkenntnis entsteht, als Wahrheit, und das Bewußtsein des Erkennenden von dem Vorliegen eines so ermittelten hohen Wahrscheinlichkeit, als die Überzeugung von der Wahrheit"²⁾

1) Bei diesem Begriff handelt es sich um die in der Rechtsprechung übliche und gemäßigte Umschreibung für den Begriff der "Wahrheit", falls "nur erkenntnistheoretische Zweifel" bestehen. So verstanden gehört der Begriff "mit an Sicherheit grenzender Wahrscheinlichkeit" in die Klasse von Fällen, deren Präzisierung einer Kontextelimination gleichkommt, d.h. dieser Begriff erweist sich letztendlich als kein Wahrscheinlichkeitsbegriff (siehe Hinweis darauf bei Stegmüller: Bd.IV, Teil A, S.132).

Eine weitere Möglichkeit besteht darin, den präzisierten Begriff "mit an Sicherheit grenzender Wahrscheinlichkeit" nicht mit dem der Wahrheit, sondern mit demjenigen der "praktischen Sicherheit" zu identifizieren (und so ihn zu eliminieren). Strenggenommen müßte man - will man der Rationalitätsforderung genügen, die nicht korrigierbare, dogmatische Wahrscheinlichkeitsbewertungen ausschließt (vgl.S.33f) - nur das logisch unmögliche Ereignis als praktisch ausgeschlossen und entsprechend das logisch notwendige Ereignis als praktisch sicher betrachten. Aber auch dann, wenn man "praktisch sicher" in einem 'schwächeren' Sinn der "Verifizierbarkeit" verwenden würde, wäre es vom Standpunkt des Richters aus, unzulässig, zu behaupten, der Sachverhalt sei verifiziert. (vgl.S.56,Fn.2).

Interessanterweise widerlegen empirische Untersuchungen (v.a. Simon & Mahan, vgl. 2.Kap.), daß es sich bei dem obigen Begriff um eine lediglich ferne theoretische Zweifel zulassende oder gar um eine performative (ritualisierte) Äußerung handelt. Versucht man nämlich die qualitativen Urteile von Einzelpersonen in eine quantitative Sprechweise zu übertragen, so ergibt sich, daß Personen darunter in der Praxis jeden beliebigen Wahrscheinlichkeitswert im Intervall von 0.5 bis 1 verstehen. (Auf diese und ähnliche Untersuchungen wird später (Kap.2,\$2) noch näher eingegangen).

Es ergibt sich somit ein Widerspruch zwischen der Meinung der theoretisierenden Rechtsprechung über den Begriff "mit an Sicherheit grenzender Wahrscheinlichkeit" und empirischen Daten, die belegen, daß hier ein Wahrscheinlichkeitswert abgeleitet wird, der zwar für die Verurteilung in einem konkreten Fall ausreicht, jedoch weit davon entfernt ist, "nur theoretische Zweifel" zu repräsentieren.

Manche Autoren schlagen für den Ausdruck "mit an Sicherheit grenzender Wahrscheinlichkeit" den numerischen Wert 0.99... vor; vgl. R.Bender/S.Röder/A.Nack, Tatsachenfeststellung vor Gericht, Bd.I (1981), S.175ff, S.194f. So gesehen handelt es sich hier nicht mehr um einen kontexteliminierbaren, sondern um einen wohldefinierten Wahrscheinlichkeitsbegriff, der allerdings auch nicht mehr "nur (erkenntnis-) theoretische Zweifel" zum Ausdruck bringt.

2) RGZ 15, S.338 (339); die Angaben sind E.Schneider, Beweis und Beweiswürdigung (1978), S.17, entnommen. Hervorhebung im Text.

So einfach, wie es das Reichsgericht in seiner Bestimmung darstellt, läßt sich das Problem allerdings nicht lösen: Denn erstens wird es sich in den allermeisten juristischen Fällen nicht - wie das Reichsgericht verharmlosend darlegt - um einen bloß theoretischen (und daher zu vernachlässigenden) Zweifel handeln, sondern um einen durch irgendwelche Tatsachen begründeten Zweifel (der Fall ist nämlich fast ausgeschlossen, daß alle Beweismittel nur zugunsten eines Sachverhalts sprechen¹⁾) - die Juristen pflegen aber stets die Urteilsverkündung mit der Wahrheitsbehauptung einzuleitung bzw. abzuschließen -; zweitens bleibt auch eine noch so hohe Wahrscheinlichkeit immer noch eine Wahrscheinlichkeit und kann nicht Wahrheit sein (vgl. in der obigen Bestimmung die Gleichsetzung des hohen Grades der Wahrscheinlichkeit mit der Wahrheit); die beiden Begriffe sind strukturell verschieden und ihre Identifizierung miteinander beruht daher auf einer Fiktion bzw. auf einem (unzulässigen) Trick; und drittens kann die durch einen solchen Trick hergestellte "Überzeugung von der Wahrheit" nicht im vom Reichsgericht intendierten Sinn als die objektive Findung der Wahrheit verstanden werden, denn die vorgegebene Überzeugung wird lediglich (sich selbst und den anderen) suggeriert, jedoch nicht tatsächlich erreicht. Eine derartige Herstellung der fiktiven Wahrheit ist somit mit dem juristischen Anspruch auf "Objektivität" unvereinbar.

Somit scheint die Feststellung eines juristischen Sachverhalts in der Rechtspraxis keine Beweis-, sondern eher eine Entscheidungsan gelegenheit zu sein. Denn der Tatbestand wird in Wirklichkeit nicht bewiesen, sondern es wird entschieden, ob man den Tatbestand für wahr oder für falsch halten soll.

Die Motivation für das Denken in der Ja-Nein-ALternative ist leicht zu erklären: Dies vereinfacht nämlich nicht nur die Entscheidungs bildung, sondern suggeriert darüber hinaus Sicherheit, Allgemein gültigkeit, Unanfechtbarkeit usw. Doch hat das natürlich seinen Preis: Die Urteilsbildung wird vergrößert, weil die Risiken einer Fehlentscheidung nicht adäquat oder überhaupt nicht bei der Ent scheidungs bildung berücksichtigt werden; und sie wird nicht im Sinne der rationalen Wahl der besten und der gerechtesten Maßnahme begriffen, denn diese Wahl ist durch die Entscheidung für einen be stimmten Sachverhalt (Tatbestand) bereits vor-entschieden.

Eine solche Haltung scheint nicht mehr zeitangemessen zu sein, nicht zuletzt deshalb, weil uns heute ein wissenschaftliches In strumentarium zur Verfügung steht, mit dessen Hilfe wir, unter Be rücksichtigung aller Arten von Unsicherheit und Risiken, die Sach lage adäquat beurteilen können²⁾

1) bzw. alle Beweismittel, die zugunsten eines anderen Sachverhalts sprechen, völlig unglaubwürdig sind.

2) Wir werden auf die Forderung "völlige Überzeugung im Bestehen des Sachverhalts" (bzw. "von der Wahrheit des Sachverhalts") noch einmal am Ende dieses Paragraphen (2.5) eingehen und die Problema tik, die sich hinter dieser Forderung verbirgt, auch von einer an deren Seite beleuchten.

Nach dieser Abschweifung in die Rechtsprechung wollen wir das Thema aus Abschnitt 2.1. wieder aufgreifen. Übertragen auf die Rechtspraxis lassen sich die Gedanken aus diesem Abschnitt kurz so zusammenfassen:

Juristische Schlüsse, in denen es um die Beurteilung eines Sachverhalts aufgrund von Informationen geht, sind induktive Schlüsse; die Bildung solcher Schlüsse gehört in den Bereich des induktiven Rasonierens. Die hier gemeinten induktiven Schlüsse sind ihrerseits Wahrscheinlichkeitsschlüsse. Will man in der Rechtspraxis adäquat und rational vorgehen, so muß man die Beurteilung juristischer Sachverhalte wahrscheinlichkeitstheoretisch rekonstruieren.

2.3. Der nächste Punkt, der uns interessiert, ist die Frage, welche Art von Wahrscheinlichkeitsschlüssen juristische Schlüsse sind. Dazu machen wir wiederum einen Vergleich. Wir stellen die Wahrscheinlichkeitsbeurteilung eines Ereignisses "Gestern nacht hat es in Mailand geregnet" der Beurteilung eines juristischen Sachverhalts "A hat B getötet" gegenüber. Unsere Informationsbasis im ersten Fall bestehe aus dem gestrigen Wetterbericht, der Regen für ganz Italien voraussagte, und der Aussage eines Bekannten, der behauptet, er sei gestern nacht in Mailand gewesen und es habe dort geregnet; im zweiten Fall sei die Informationsbasis ein Motiv und die Aussage eines Zeugen, die besagt, daß A vom Zeugen am Tatort zur fraglichen Zeit gesehen wurde. Nun, wie glaubhaft ist die erste Hypothese? Das wird sicher davon abhängen, für wie zuverlässig wir den Wetterbericht halten, wie zuverlässig der Bekannte in unseren Augen ist, ob wir der Ansicht sind, daß "für ganz Italien" notwendigerweise "Mailand" einschließt oder lediglich eine Abkürzung für einen viel komplexeren Zusammenhang (etwa wie "für ganz Italien, aber es kann der Fall sein, daß einige Orte ausgenommen sind") darstellt und von sonstigen Allgemeinkenntnissen und Erfahrungen. Angenommen, wir sind der Ansicht, daß Wetterberichte in der Mehrzahl aller Fälle stimmen, daß unser Bekannte einen sehr zuverlässigen Eindruck macht und daß unsere sonstige Erfahrung bestätigt, daß es in Mailand zu dieser Jahreszeit oft regnet. Dies alles beeinflußt unsere intuitive Beurteilung des Ereignisses und macht seine Glaubwürdigkeit in unseren Augen bestimmt größer als wenn uns keine oder weniger zuverlässige Informationen zur Verfügung stünden.

Analog im zweiten Fall: Es wird viel davon abhängen, wie handfest das Motiv des Angeklagten ist, wie zuverlässig die Zeugenaussage ist, und auch von sonstigen Informationen, die vielleicht nachträglich beschafft werden konnten. Angenommen, das Motiv ist in den Augen des Richters handfest und der Zeuge scheint ihm sehr glaubwürdig, der Angeklagte aber hat ein nicht minder zuverlässiges Alibi (d.h. die Informationsbasis sei durch eine neue Information erweitert). Es ist klar, daß die Schlußfolgerung aufgrund der erweiterten Informationsbasis anders ausfallen wird als aufgrund der ursprünglichen. (Vielleicht gelingt es später aufgrund zusätzlicher Informationen, zu beweisen, daß der Zeuge doch nicht so zuverlässig ist, wie es am Anfang schien, oder daß der Angeklagte nicht der einzige war, der das Motiv hatte usw., so daß die Beurteilung des Sachverhalts schließlich doch zugunsten des Angeklagten ausfällt).

Wie wir sehen, besteht wiederum kein prinzipieller Unterschied zwischen der Beurteilung eines alltäglichen Ereignisses und der eines juristischen. In Abschnitt 1.2.1. haben wir festgestellt, daß der Beurteilung solcher Ereignisse ein vorwissenschaftlicher, intuitiver Begriff der Wahrscheinlichkeit zugrunde liegt, der durch den subjektiven Wahrscheinlichkeitsbegriff zu präzisieren ist.

Es besteht also kein Grund, wieso wir das Sprechen über juristische Sachverhalte nicht im Rahmen der subjektiven Wahrscheinlichkeitstheorie präzisieren und die Beurteilung juristischer Sachverhaltshypothesen nicht aufgrund des subjektiven Wahrscheinlichkeitsbegriffs vornehmen können oder sollen.

So verstanden drücken die präzisierten juristischen Wahrscheinlichkeitsurteile nichts anderes aus als die rationale Überzeugung des Richters vom (oder den rationalen Glauben des Richters an das) Bestehen bestimmter Sachverhalte aufgrund der zur Verfügung stehenden Informationen und Erfahrungen.

Ein weiterer Vorteil der Rekonstruktion juristischer Schlüsse aufgrund des subjektiven Wahrscheinlichkeitsbegriffs ist der, daß nicht nur die Sachverhalts-, sondern auch die Entscheidungsfrage rational angegangen werden kann. Es sind nämlich subjektive Wahrscheinlichkeitsbewertungen, die unserem Handeln bzw. Entscheiden zugrunde lie-

gen. Somit ist es naheliegend, die juristische Entscheidungsbildung, die ja auch von der subjektiven Wahrscheinlichkeitsbewertung des Sachverhalts ausgeht, im Rahmen einer rationalen Entscheidungstheorie zu rekonstruieren.

Immer dann, wenn uns Gewißheit fehlt, wir jedoch handeln oder entscheiden müssen, verwenden wir intuitiv eine Bewertung mit Wahrscheinlichkeiten. Es besteht also ein enger Zusammenhang zwischen subjektiven Wahrscheinlichkeitsbewertungen und Entscheidungen. Dieser wird daher insbesondere in der rationalen Entscheidungstheorie, die eine Präzisierung, ein rationales Modell dieses Zusammenhangs bildet, explizit dargelegt. Wie bekannt, ist neben dem subjektiven Wahrscheinlichkeitsbegriff der subjektive Nutzenbegriff grundlegend für diese Theorie. In der rationalen Entscheidungstheorie will man einen allgemeinen Rahmen entwickeln, innerhalb dessen man von konsistenten Glaubensannahmen und konsistenten Nutzenbewertungen von Konsequenzen zu rationalen Entscheidungen gelangt.

Wir können somit nicht umhin, zu versuchen, Bereiche, in denen praktische Entscheidungen in Verbindung mit subjektiven Bewertungen von unsicheren Ereignissen oder Sachverhalten getroffen werden, entscheidungstheoretisch zu rekonstruieren. Es scheint daher selbstverständlich, auch die strafrechtliche Praxis, insofern sie von subjektiven Bewertungen von unsicheren Sachverhalten ausgeht und wo letztlich praktische Entscheidungen getroffen werden, in diesem Rahmen aufzubauen. (Vgl. den II. Teil dieser Arbeit).

2.4. Juristische Urteile sind Urteile, denen ein intuitiver Wahrscheinlichkeitsbegriff zugrunde liegt, der durch den subjektiven Wahrscheinlichkeitsbegriff präzisiert werden kann. Um so erstaunlicher ist es, daß diese Auffassung in der juristischen Fachliteratur offensichtlich nicht geteilt wird¹⁾. Wenn man sich über-

1) Eine der wenigen erfreulichen Ausnahmen ist das Buch von E.L.Nell, Wahrscheinlichkeitsurteile in juristischen Entscheidungen (1983). Nell ist meines Erachtens der erste im deutschsprachigen Raum, der in ausführlicher (allerdings informeller) Weise darauf hinweist, daß der juristische Wahrscheinlichkeitsbegriff subjektiver Art ist. Siehe aber auch A.Evers, Begriff und Bedeutung der Wahrscheinlichkeit für die richterliche Beweiswürdigung (1979).

Auch im anglo-amerikanischen Sprachraum sind die "Subjektivisten" unter den Juristen, die sich mit Wahrscheinlichkeitsargumentationen befassen, weniger zahlreich als die "Objektivisten". Zu den "Subjektivisten" gehören u.a. Cullison, Kaye, McBaine, Schum, Birmingham (Auf die anglo-amerikanischen Beiträge zur Diskussion der Anwendung der Wahrscheinlichkeitstheorie auf die Jurisprudenz wird im 2.Kap. dieses Teils näher eingegangen).

Vgl. D.Kaye, The Law of Probability and the Law of the Land, in: 47 The Uni. of Chicago Law Rev. (1979), S.34-56; S.48: "...there is nothing paradoxical or counterintuitive in thinking of legal factfinding in terms of the mathematics of probability and that the subjective definition of probability provides a meaningful interpretation of the probabilities involved in the legal process"; D.A.Schum, The Case Against Pascal, in: 77 Michigan Law Rev. (1979), S.446-483; S.477: "...I believe that the only extant interpretation which does merit serious consideration in juridical inference is the personalistic or subjective approach..."; vgl. auch Wigmore, der vom juristischen Beweiss als "the degree of strength of belief of a fact-in-issue pro-

haupt mit der Anwendung der Wahrscheinlichkeitstheorie in der Rechtspraxis befaßt¹⁾ – und die steigende Tendenz dazu ist seit ca. 25 Jahren zu beobachten –, so nimmt man größtenteils unbegründet an, daß es sich dabei um den objektiven Wahrscheinlichkeitsbegriff im Sinne der relativen Häufigkeit handelt.

Als Beleg dafür kann fast die gesamte zum Wahrscheinlichkeitsbegriff erschienene juristische Literatur angesehen werden.

Im deutschen Sprachraum vertreten in besonders krasser Form Weitnauer, Schreiber, Musielak und Maassen die Häufigkeitsauffassung²⁾.

So definiert Schreiber (vgl. S. 25f) den sog. "Beweiswert eines Beweismittels" folgendermaßen: Der Beweiswert ist "die Wahrscheinlichkeit der Wahrheit der Aussage, diese wiederum... das Verhältnis der Anzahl der wahren Aussagen zu der Anzahl der wahren und falschen Aussagen" (Interessant wäre es zu wissen, wie Schreiber mit dem Problem der sich ständig verändernden Bezugsklasse der wahren und falschen Aussagen und dementsprechend der sich ständig ändernden relativen Häufigkeit der "wahren" Aussagen in der Gesamtheit fertig wird. Die Antwort darauf bleibt Schreiber uns schuldig)³⁾.

Vgl. Weitnauer (S. 6): "Eine Wahrscheinlichkeitsaussage läßt sich sinnvollerweise nur in bezug auf ein Kollektiv, eine statistische Masse, nicht aber in bezug auf einen einzelnen Fall machen". Allerdings ist dazu folgendes anzumerken: erstens gibt Weitnauer selbst zu, von der subjektiven Wahrscheinlichkeitstheorie wenig zu verstehen (vgl. S. 4: "Ich gestehe, daß ich nicht in der Lage war, solchen Gedankengängen (gemeint sind subjektivistische – Anm. von mir) zu folgen, zumal sie auch noch mit den Methoden der symbolischen Logik und im Gewande der

duced by evidence on the mind of the jury" spricht. (J. Wigmore, Evidence § 2498a).

Es sei darauf hingewiesen, daß es sich weder bei der "subjektivistischen" noch der "objektivistischen" Richtung in der modernen Jurisprudenz um konsequent durchdachte und theoretisch aufgebaute Positionen handelt, sondern lediglich um zum größten Teil vage Ansichten, Meinungen und Mutmaßungen zur Natur des juristischen Wahrscheinlichkeitsbegriffs.

1) Denn die etablierte und maßgebende Richtung in der Rechtspraxis geht noch immer davon aus, daß die Wahrheit feststellbar sei, bzw. – falls "nur theoretische Zweifel bestehen" – "mit an Sicherheit grenzender Wahrscheinlichkeit" feststellbar sei. In jedem Fall müsse daher der Richter "vom Bestehen des Sachverhalts völlig überzeugt" sein. Die Annahme eines Wahrscheinlichkeitsbegriffs sei daher überflüssig. Vgl. E. Döring, Die Erforschung des Sachverhalts im Prozeß (1964), S. 6ff und S. 447ff.

2) H. Weitnauer, Wahrscheinlichkeit und Tatsachenfeststellung, in: Karlsruher Forum (1966); R. Schreiber, Theorie des Beweiswerts für Beweismittel im Zivilprozeß (1968); Musielak, Die Grundlagen der Beweislast im Zivilprozeß (1975); B. U. Maassen, Beweismaßprobleme im Schadensersatzprozeß (1975).

3) Zur ausführlichen Kritik an Schreiber vgl. S. 103ff dieser Arbeit.

höheren Mathematik vorgetragen werden. Ich wage deshalb auch nicht zu beurteilen, ob darin Anregungen und Ansätze stecken, die in unserem Bereich (gemeint ist die Jurisprudenz - Anm. von mir) weiterführen könnten; da ich das aber für möglich halte, kann ich deshalb nur der Hoffnung Ausdruck geben, daß sich bald jemand finden wird, der in der Lage ist, diesen Gedankengängen vom Boden unserer Wissenschaft aus nachzugehen"); und zweitens beschränkt er die Anwendung des statistischen Wahrscheinlichkeitsbegriffs auf eine enge Klasse von zivilrechtlichen Fällen, in denen es z.B. um die Ermittlung einer durchschnittlichen Lebenserwartung (wie bei der Kapitalisierung einer Rente) oder um die Ermittlung der Wahrscheinlichkeit eines Schadens usw. geht und in denen es zum Teil auf den Vergleich mit anderen "ähnlich gelagerten" Fällen ankommt, so daß man hier (bei der Beurteilung von Beweismitteln) einen Quasi-Häufigkeitsbegriff heranziehen kann¹⁾.

Vgl. Musielak (S.112): "Der Wahrscheinlichkeitsbegriff der Wahrscheinlichkeitstheorie, der für die Rechtswissenschaft Gültigkeit hat, läßt sich aber in objektiv bestimmbare Grade²⁾ einteilen und an ihm kann es keineswegs liegen, wenn der Aufklärungseifer erlahmt und von vagen Vorstellungen bei der Tatsachenwürdigung ausgegangen wird. Wahrscheinlichkeit in diesem Sinn bedeutet der objektive Befund, der sich für die Richtigkeit oder Unrichtigkeit einer bestimmten Hypothese ergibt; seine Grundlage bildet die Feststellung der relativen Häufigkeit bestimmter Erscheinungen". Da Musielak keine formale Begründung seiner Behauptung liefert und darüber hinaus eine ziemlich ungenaue und teilweise unrichtige inhaltliche Bestimmung des Wahrscheinlichkeitsbegriffs angibt (vgl. seine Äußerung bzgl. der Einzelfallwahrscheinlichkeit, S.112: "Wenn von der Wahrscheinlichkeit im Einzelfall²⁾ gesprochen wird, muß man sich darüber im klaren sein, daß in dieser Aussage eine aus Gründen sprachlicher Abkürzung übliche Ungenauigkeit liegt. Denn eine Wahrscheinlichkeitsaussage läßt sich immer nur bezüglich einer statistischen Masse, eines Kollektivs, nicht dagegen bezüglich eines Einzelfalls machen"), wäre in seinem Fall die Annahme gar nicht fehl am Platze, daß Musielak selbst der von ihm angesprochenen Krankheit "des Erlahmens des Aufklärungseifers" erlegen und von "vagen Vorstellungen" in bezug auf den Wahrscheinlichkeitsbegriff ausgegangen ist.

Auch Maassen vertritt die Häufigkeitsauffassung (S.6): "Dieser Wahrscheinlichkeitsbegriff, der auf der sogenannten Häufigkeitstheorie der Wahrscheinlichkeit beruht, ist ohne weiteres auch auf die juristische Tatsachenfeststellung anwendbar". Da Maassen noch darüber hinaus ein nicht adäquates entscheidungstheoretisches Modell angibt, wobei der Zusammenhang zwischen den als relative Häufigkeiten aufgefaßten Wahrscheinlichkeiten des Sachverhalts und sonstigen subjektiven Bewertungen nicht recht verständlich wird, kann man auch ihm vorwerfen, daß er sich nicht nur in bezug auf den Wahrscheinlichkeitsbegriff, sondern auch in bezug auf die Entscheidungstheorie nicht genug Klarheit verschafft hat (wir werden in II. Teil dieser Arbeit das Entscheidungsmodell von Maassen näher untersuchen).

1) Weitnauer geht allerdings davon aus, daß es sich hier nicht um einen Quasi-Häufigkeitsbegriff handelt, sondern um einen wohldefinierten mathematischen Häufigkeitsbegriff; vgl. S.18: "Um eine Wahrscheinlichkeitsaussage zu ermöglichen, muß der Fall als Element eines Kollektivs betrachtet werden können, bei dem dann Schäden gewissen Umfangs oder gewissen Verlaufs mit gewisser relativer Häufigkeit auftreten". Dazu ist anzumerken, daß in den allermeisten Fällen weder das zugrundezulegende Kollektiv genau beschreibbar noch die relative Häufigkeit (auch in formaler Hinsicht, für einen neuen Fall) feststellbar ist. Vgl. dazu S.68f dieser Arbeit.

2) Hervorhebung im Text.

Äußerungen zum objektiven Wahrscheinlichkeitsbegriff findet man auch bei G.Kegel, R.Motsch, E.Schneider und W.Berg¹⁾
 So schreibt Schneider in bezug auf den Vaterschaftsnachweis (S.20): "Die Statistik ersetzt also die Überzeugung"²⁾, wenn man es überspitzt ausdrücken will".
 Eine ähnliche Auffassung vertritt Berg (S.74): "Bei anderer Ausgestaltung des gesetzlichen Tatbestandes kann unter Umständen eine hohe statistische Wahrscheinlichkeit oder eine exakte naturwissenschaftliche Berechnung den subjektiven Überzeugungsvorgang praktisch ersetzen".

Nicht klar ist die Auffassung Hippels bzgl. des Wahrscheinlichkeitsbegriffs³⁾
 Hippel nimmt (vgl. S.26f) für verschiedene Bereiche wie der Rückfallprognose bei der Entscheidung über Strafaussetzung zur Bewährung, der Kriminalprognose bei der bedingten Entlassung, der Behandlungsprognose etc. verschiedene, wie er schreibt, "undefinierte", Wahrscheinlichkeitsbegriffe an. Zur Begründung, wieso der/die juristische/n Wahrscheinlichkeitsbegriff/e "undefiniert" bleiben soll/en, führt Hippel folgendes, nicht überzeugendes, Argument an: Die "Einführung eines bestimmten vorrechtlichen Wahrscheinlichkeitsbegriffs muß schon daran scheitern, da es bereits innerhalb der Wahrscheinlichkeitstheorie und der allgemeinen Wissenschaftslehre keinen einheitlichen Wahrscheinlichkeitsbegriff gibt, so daß sich schon deshalb jeglicher juridischer Apriorismus verbietet"(S.26).

Ähnliches gilt auch für Gottwald⁴⁾, der zwar eine "Erlebniswahrscheinlichkeit des Alltags (eine Art subjektiver Wahrscheinlichkeit) als Beurteilungskriterium juristischer Sachverhalte annimmt"(S.191, 206), sich jedoch weigert, diesen Begriff im subjektivistischen Sinn zu verstehen (S.193).

Eine einmalige Position nimmt Greger⁵⁾ ein: Er lehnt sowohl den subjektiven wie auch den objektiven Wahrscheinlichkeitsbegriff ab (S.49ff) und nimmt irrtümlicherweise an, daß der logische Wahrscheinlichkeitsbegriff eine Präzisierung des juristischen Wahrscheinlichkeitsbegriffs darstellt (S.55f), ohne sich der Tatsache bewußt zu sein, daß der logische Wahrscheinlichkeitsbegriff ein subjektiver Wahrscheinlichkeitsbegriff ist, für den außer den Kolmogoroff-Axiomen eine Reihe zusätzlicher Rationalitätskriterien gilt, die für die Rechtspraxis unangebracht sind⁶⁾.

1) G.Kegel, Der Individualanscheinsbeweis und die Verteilung der Beweislast nach Überwiegender Wahrscheinlichkeit, in: Festgabe für H.Kronstein (1967), S.321ff (darauf wird im 2.Kap. näher eingegangen); R.Motsch, Vom Prozeß als Beweis zum Überwiegensprinzip, in: Gedächtnisschrift für Jürgen Rüdiger (1978), S.334ff; E.Schneider, Beweis und Beweiswürdigung unter besonderer Berücksichtigung des Zivilprozesses (1978); W.Berg, Die verwaltungsrechtliche Entscheidung bei ungewissem Sachverhalt (1980).

2) Hervorhebung im Text.

3) R.v.Hippel, Gefahrenurteile und Prognoseentscheidungen in der Strafrechtspraxis (1972).

4) P.Gottwald, Schadenszurechnung und Schadensschätzung (1979).

5) R.Greger, Beweis und Wahrscheinlichkeit (1978).

6) Zum Begriff der logischen Wahrscheinlichkeit vgl. Abschnitt 1.2.3, S.50ff dieser Arbeit.

Wenig Verständnis für die subjektive Wahrscheinlichkeitstheorie zeigen auch Bender/Röder/Nack¹⁾ auch dann, wenn sie (wie auf S.389) schreiben, daß es sich im Prozeß "um einen sehr hohen Grad subjektiver Wahrscheinlichkeit"²⁾ handelt. Denn der Leser erfährt, daß es den Autoren tatsächlich nicht um die Ermittlung subjektiver Glaubensgrade geht, sondern um eine solche der "Häufigkeit"²⁾ des Auftretens einer Indizienkombination (oder eines einzelnen Indizes) zusammen mit der Haupttatsache (d.h. der Hypothese - Anm. von mir) und ohne diese" (S.188). Auch das Bayes-Theorem wird von Bender/Röder/Nack entsprechend einer Häufigkeitsauffassung interpretiert; vgl. S.191:
 "Auch das Theorem von Bayes baut auf der - anzunehmenden - Anfangs ("Ausgangs")-wahrscheinlichkeit auf, nämlich:

Anzahl der Fälle,

in welchen die Haupttatsache vorliegt

Anzahl aller (vergleichbaren) Fälle,

in welchen die Haupttatsache vorliegt

oder auch nicht vorliegt...

$$\frac{n(A)}{n(A \cup \bar{A})} = P(A)''$$

Bender/Röder/Nack weisen übrigens (was wichtig ist) darauf hin, daß man nicht nur die Unsicherheit bzgl. der Schlußfolgerung von Daten auf den Sachverhalt, sondern auch eine solche bzgl. der Daten selbst berücksichtigen muß. Leider wenden sie in allen Beispielen, in denen die Unsicherheit von Daten berücksichtigt werden soll, die Multiplikationsregel an: d.h. sie multiplizieren die Wahrscheinlichkeit des Sachverhalts aufgrund von Daten mit der Wahrscheinlichkeit für die Daten. Dies ist aber nur dann korrekt, wenn die Wahrscheinlichkeit für die Hypothese aufgrund der Negation der Daten 0 ist (dies trifft tatsächlich in den meisten von Bender/Röder/Nack konstruierten Beispielen, wie etwa auf S.203, zu, wird aber nicht von diesen explizit vorausgesetzt (dies ist aber für eine korrekte Anwendung unerlässlich)). Anstelle des Multiplikationsprinzips muß man korrekterweise die Jeffrey-Formel³⁾ anwenden (die sich zu der von Bender/Röder/Nack benutzen Multiplikationsregel verkürzt, falls die oben angesprochene Voraussetzung erfüllt ist).

Wir werden uns im nächsten Kapitel die Diskussion des Wahrscheinlichkeitsbegriffs in der juristischen (vorwiegend anglo-amerikanischen) Fachliteratur näher ansehen. Eine gute Darstellung verschiedener Auffassungen deutscher Autoren kann der Leser bei Nell⁴⁾ finden⁵⁾

1) R.Bender/S.Röder/A.Nack, Tatsachenfeststellung vor Gericht, Bd.I (1981), S.173ff.

2) Hervorhebung von mir.

3) Vgl. S.34f dieser Arbeit.

4) Vgl. S.62, Fn.1 dieser Arbeit.

5) Einen Überblick über die deutsche juristische Fachliteratur zum Wahrscheinlichkeitsbegriff gibt auch A.Evers in: Begriff und Bedeutung der Wahrscheinlichkeit für die richterliche Beweiswürdigung (1979) an.

Die Annahme, daß es sich bei dem juristischen Wahrscheinlichkeitsbegriff um den Begriff der objektiven Wahrscheinlichkeit (bzw. den der relativen Häufigkeit) handelt, kann nur auf Unkenntnis der Natur des objektiven Wahrscheinlichkeitsbegriffs bzw. auf der naiven Interpretation von 'objektiv' beruhen (wie die Ablehnung des subjektiven Wahrscheinlichkeitsbegriffs durch manchen Juristen sich auf eine naive Auffassung von 'subjektiv' als 'willkürlich' stützt).

Die Motivation, die dem Denkenwollen in objektiven Wahrscheinlichkeiten zugrunde liegt, ist vielleicht gar nicht so schwer zu durchschauen. Man könnte sie wie folgt beschreiben: Auch ein Jurist will - in Anlehnung an die Naturwissenschaften - objektiv, d.h. unabhängig von subjektiven Meinungen und Einschätzungen, urteilen können, und wenn ihm schon die Wahrheit (vom Bestehen des Sachverhalts) nicht zugänglich ist, so möchte er wenigstens mit einem Wahrscheinlichkeitsbegriff arbeiten, der etwas mit Objekten bzw. mit Tatsachen selbst zu tun hat, der etwas über die Objekte und Tatsachen selbst aussagt, und nicht bloß Ausdruck einer persönlichen Erwartungshaltung ist. Und da, seiner Meinung nach, der objektive statistische Wahrscheinlichkeitsbegriff dies alles garantiert, versucht er (trotz der Schwierigkeiten, die hier auftreten und die ihm inzwischen partiell bekannt sind¹⁾) unter gewissen Verdrehungen den statistischen Begriff zurechtzubiegen, damit daraus ein für die Rechtspraxis 'brauchbarer' Wahrscheinlichkeitsbegriff entsteht²⁾

Diese Auffassung verkennt nicht nur, daß die Jurisprudenz keine exakte Wissenschaft wie die Physik ist, in der es um die wahrscheinlichkeitstheoretische Beurteilung von 'gleichartigen', prinzipiell beliebig oft wiederholbaren Vorgängen geht (während jene beim besten Willen nicht als solche rekonstruierbar ist), sondern mißdeutet auch völlig die Besonderheit der juristischen Argumentationssituation: In dieser, wie in jedem anderen Bereich des induktiven (intuitiven, alltäglichen) Rasonieren, geht es gerade um subjektive

1) So weist man in der juristischen Literatur zum Wahrscheinlichkeitsbegriff gewöhnlich auf die Problematik des Sprechens von Einzelfällen innerhalb der objektivistischen Theorie hin, erwähnt aber nirgend die Schwierigkeiten bzgl. der Definition des objektiven Wahrscheinlichkeitsbegriffs.

2) Vgl. den auf S. 64 dieser Arbeit angesprochenen Versuch von Musielak, den Häufigkeitsbegriff zu rechtfertigen, indem er fälschlich behauptet: "Wenn von der Wahrscheinlichkeit im Einzelfall gesprochen wird, muß man sich darüber im klaren sein, daß in dieser Aussage eine aus Gründen sprachlicher Abkürzungen übliche Ungenauigkeit liegt" (S. 112); Hervorhebung im Text.

Meinungen und Einschätzungen; diese sind präzisierungsbedürftig und können präzisiert werden; vor allem sollen aber derart präzierte Meinungen und Einschätzungen bezüglich juristischer Sachverhalte einer rationalen Entscheidungsbildung zugrunde gelegt werden. Um Rationalität, nicht um die in der Statistik gemeinte Objektivität muß es sich also in der Rechtspraxis handeln.

Man könnte meinen, daß auch juristische Sachverhalte, wie Mord, Diebstahl etc. einen Massencharakter haben. Wieso sollte man dann nicht statistisch vorgehen?

Der Vergleich mit einem statistischen Vorgang wie dem Ziehen einer schwarzen Kugel aus einer Urne mit 10 Kugeln, von denen nur 2 schwarz sind, zeigt aber, daß diese Annahme falsch ist.

In der obigen Situation ist die Wahrscheinlichkeit für das Ziehen einer schwarzen Kugel gleich $1/5$. Sowohl die Bezugsklasse, d.i. die Menge aller Kugeln in der Urne wie auch die Anzahl der Kugeln mit der in Frage kommenden Eigenschaft sind fest angegeben. Es hat daher einen Sinn, die objektive Wahrscheinlichkeit für das Ziehen einer schwarzen Kugel mit der relativen Häufigkeit des Auftretens einer schwarzen Kugel in der Urne gleich zu setzen.

Wissen wir aber weder, wieviele Kugeln in der Urne, noch wieviele davon schwarz sind, so können wir gar nicht die relative Häufigkeit für das Auftreten der schwarzen Kugeln in der Urne angeben. Angenommen, eine bei dem Versuch anwesende Person, die über dieses Verhältnis offensichtlich Bescheid weiß, gibt uns einen Hinweis, mit dessen Hilfe wir gewisse Annahmen bezüglich der Wahrscheinlichkeit für das Ziehen einer schwarzen Kugel machen können. In diesem Fall handelt es sich aber nicht mehr um die Bestimmung der objektiven Wahrscheinlichkeit im Sinn der relativen Häufigkeit für das Ziehen einer schwarzen Kugel, sondern um unsere subjektive Erwartung (die in Form einer subjektiven Wahrscheinlichkeitsannahme ihren Ausdruck findet), die nicht stärker sein kann als die Glaubwürdigkeit der beim Versuch anwesenden Person für uns.

Letzteres ist mit der Situation in der Rechtspraxis vergleichbar. Der Normalfall wird hier sein, daß wir weder über die Bezugsklasse genaue Aussagen machen (die Bezugsklasse ist hier in der Regel viel komplexer, unüberschaubarer und umfaßt viel mehr Merkmale als die oben angesprochene Bezugsklasse, die nur 10 wohldefinierte Elemente enthält) noch die Anzahl der Träger mit den in Frage kommenden Eigenschaften bestimmen werden können. Es hat daher keinen Sinn, in Bezug auf juristische Sachverhalte wie Mord, Diebstahl etc. von der objektiven Wahrscheinlichkeit dafür, daß der Angeklagte der Mörder ist, bzw. daß der Angeklagte den Diebstahl begangen hat etc. zu reden, geschweige denn, diese der relativen Häufigkeit von Mördern oder Dieben unter Personen mit bestimmten Eigenschaften gleichzusetzen.

Wir können noch einen weiteren Grund angeben, wieso es mit der statistischen Bestimmung nicht funktioniert. Wir nehmen vereinfacht an, daß bei 10 verschiedenen Mordfällen in einem bestimmten Zeitraum jeweils ein Fingerabdruck festgestellt wurde. Ferner nehmen wir an, daß in jedem Fall die Person, die den Fingerabdruck am Tatort hinterlassen hat, ausfindig gemacht, aber nur in zwei der Fälle sie auch als Mörder identifiziert wurde (wir wissen ja, daß auch dann, wenn eine Person als Träger der am Tatort gefundenen Fingerabdrücke in Frage kommt, diese nicht unbedingt der Mörder sein muß). Die relative Häufigkeit der identifizierten Mörder in der Gesamtheit dieser 10 Fälle (mit der Bezugsklasse "die am Tatort hinterlassenen Fingerabdrücke") ist somit $1/5$.

Angenommen, wir haben einen neuen Fall, in welchem wiederum die verdächtige Person die gleichen Fingerabdrücke hat wie diejenigen, die am Tatort gefunden wurden. Dann können wir aber nicht die objektive Wahrscheinlichkeit dafür, daß diese Person auch die Tat begangen hat, gleich $1/5$ setzen, denn die Bezugsklasse umfaßt jetzt 11 Elemente (und nicht 10, in bezug auf welche die relative Häufigkeit festgestellt wurde); für diesen Fall ist die Wahrscheinlichkeit aber noch nicht festgestellt.

(Wären wir im vorliegenden Fall genauer, so würden wir nicht einmal die Wahrscheinlichkeit von $1/5$ als relative Häufigkeit auffassen. Denn bei der obigen Feststellung, daß unter verdächtigen Personen, die ihren Fingerabdruck am Tatort (in 10 Fällen) hinterlassen haben, nur zwei Mörder sind, handelt es sich nicht um die wahre Tatsachenbeschreibung, sondern lediglich um das Für-wahr-Halten des beurteilenden Richters (ob diese zwei Personen tatsächlich Mörder waren, bzw. ob es in den 10 Fällen vielleicht doch eine ganz andere Verteilung hinsichtlich des Merkmals "Mörder" unter den verdächtigen Personen gibt, läßt sich nicht objektiv nachprüfen). Gerade hier sehen wir aber, wie subjektive Einschätzungen und Meinungen (in Form subjektiver Wahrscheinlichkeitsannahmen) die Beurteilung eines juristischen Sachverhalts bestimmen).

Bei der obigen Darstellung haben wir bewußt davon Abstand genommen, auf die vielerorts in dieser Arbeit angesprochene prinzipielle Schwierigkeit der Gleichsetzung der Wahrscheinlichkeit im Einzelfall mit der relativen Häufigkeit des Auftretens bestimmter Merkmale in einer Gesamtheit von Fällen näher einzugehen bzw. hinzuweisen.

Was "Wahrscheinlichkeit als eine Eigenschaft oder Disposition von Objekten bzw. (physikalischer) Vorgänge" angeht, so irrt sich der Jurist auch hier, wenn er annimmt, daß dieser Ausdruck in bezug auf juristische Sachverhalte ebenfalls angemessen sei.

Hat es einen Sinn, hinsichtlich statistischer Phänomene (wie dem Werfen eines Würfels, dem Ziehen einer schwarzen Kugel aus der Urne, die 10 Kugeln, darunter 2 schwarze, enthält, usw.) von der objektiven Wahrscheinlichkeit als einer Eigenschaft oder Disposition zu sprechen (die Wahrscheinlichkeit $1/6$ für das Werfen der Zahl "1" stellt dann eine Disposition des Würfels selbst dar, wie die Wahrscheinlichkeit $1/5$ eine Disposition des Urnenversuchs darstellt), so ist es sinnlos, in bezug auf juristische Sachverhalte wie Mord, Totschlag, Diebstahl usw. von irgendwelchen Wahrscheinlichkeiten als Eigenschaften dieser Sachverhalte zu reden: Denn entweder hat A B getötet, A den Diebstahl begangen usw., oder er hat dies nicht getan. Der Sachverhalt besteht oder besteht nicht, von der objektiven Wahrscheinlichkeit als einer Eigenschaft des Sachverhalts kann dabei keine Rede sein. Da uns aber die Wahrheit in den allermeisten juristischen Fällen verborgen bleibt (gemeint sind hier solche Fälle, wo man zwar über das Eintreten bestimmter Ereignisse befinden soll, selbst jedoch nicht dabei war, als diese sich ereigneten, und daher diese aus irgendwelchen Informationen erschließen muß; dies ist die typische Situation vor Gericht), müssen wir eben sehen, wie wir unsere Überzeugungshaltung gegenüber dem Eintreten von juristischen Ereignissen aufgrund der zur Verfügung stehenden Erfahrungen und Informationen modifizieren.

Somit ist der einzig adäquate Wahrscheinlichkeitsbegriff bei der Beurteilung juristischer Sachverhalte der subjektive Wahrscheinlichkeitsbegriff.

2.5. Obwohl inzwischen klar sein dürfte, daß die Beurteilung juristischer Sachverhalte durch den subjektiven Wahrscheinlichkeitsbegriff präzisierbar ist, kann ein traditioneller Jurist bei der Identifizierung seines Überzeugungsbegriffs mit dem subjektiven Wahrscheinlichkeitsbegriff trotzdem Schwierigkeiten haben. Dies soll im folgenden näher untersucht werden.

Ein Richter spricht auch dann von seiner "völligen Überzeugung vom Bestehen des Sachverhalts", wenn im Laufe der Beweiswürdigung empirische Befunde, wie z.B. Zeugenaussagen, vorgelegt werden, die zwar das Gegenteil des Sachverhalts stützen, dem Richter allerdings weniger glaubwürdig erscheinen als empirische Befunde, die zugunsten des Sachverhalts sprechen (bzw. wenn die Schlußfolgerung auf den Sachverhalt aufgrund der den Sachverhalt stützenden Befunde dem Richter glaubwürdiger erscheint als die Schlußfolgerung auf die Negation des Sachverhalts aufgrund der die Negation stützenden Befunde). Hier erreicht der Richter seine "völlige Überzeugung" offensichtlich nicht nur, indem er die "sehr hohe Wahrscheinlichkeit" als "Wahrheit" identifiziert, sondern auch, indem er die weniger glaubwürdigen, das Gegenteil stützenden Daten ausschaltet bzw. vernachlässigt.

Aus wahrscheinlichkeitstheoretischer Sicht verletzt der Richter in diesem Fall die methodologische Forderung, nach der das gesamte für den Fall vorhandenen relevante Datum - und dazu gehören auch die vom Richter ausgeschlossenen empirischen Befunde - bei der Beurteilung des Sachverhalts berücksichtigt werden muß.

In welche Schwierigkeiten wir geraten, wenn wir nicht das relevante Gesamtdatum berücksichtigen, soll an einem Beispiel erläutert werden!)

Nehmen wir an, daß Y ein Schwede ist und wir die Hypothese beurteilen wollen, daß Y römisch-katolisch ist.

Die Wahrscheinlichkeit, daß Y römisch-katolisch ist unter der Voraussetzung, daß er ein Schwede ist, ist offensichtlich sehr gering (denn fast alle Schweden sind protestantisch). Aus "juristischer Sicht" müßten wir diese Hypothese somit zurückweisen und das Gegenteil davon, nämlich, daß Y nicht römisch-katolisch ist, als "wahr" annehmen. Nehmen wir weiter an, daß wir erfahren, daß Y im vorigen Jahr nach Lourdes pilgerte. Die Wahrscheinlichkeit aber, daß jemand römisch-katolisch ist, wenn er nach Lourdes pilgert, ist sehr hoch. Es ist also beinahe sicher, daß Y, bedingt durch diese Information, römisch-katolisch ist.

1) Dieses und ähnliche Beispiele findet der Leser bei W. Stegmüller, Bd. I, S. 631ff, der sich mit dem Problem der Nichtberücksichtigung des relevanten Gesamtdatums bzw. mit der Problematik des sog. "statistischen Syllogismus" ausführlich befaßt.

Aus "juristischer Sicht" müßte man dann diese Hypothese akzeptieren und diesmal die Negation davon, nämlich, daß Y nicht römisch-katolisch ist, zurückweisen. Wir geraten also in ein richtiges Dilemma, wenn wir die "juristische Sicht" beibehalten.

Aus diesem Dilemma gibt es einen - von Carnap¹⁾ vorgeschlagenen - Ausweg, der in der Berücksichtigung des Gesamtdatums besteht. Die Forderung des Gesamtdatums besagt, daß bei der Anwendung des Wahrscheinlichkeitsbegriffs auf bestimmte Situationen das gesamte verfügbare Erfahrungswissen als Beurteilungsgrundlage zu nehmen ist²⁾.

In unserem Beispiel müssen wir somit bei der probabilistischen Beurteilung der Hypothese, daß Y römisch-katolisch ist, das Gesamtdatum: "Y ist ein Schwede und Y pilgerte im vorigen Jahr nach Lourdes" berücksichtigen. Die Wahrscheinlichkeit der Hypothese, bedingt durch das Gesamtdatum, ist dann offensichtlich weder so gering wie eine solche aufgrund des ersten Datums noch so hoch wie ein solches aufgrund des zweiten Datums³⁾⁴⁾.

Macht man sich die Problematik bewußt, die aus einer rechtspolitisch motivierten Vernachlässigung bestimmter Schlußfolgerungen und Beweise seitens der Rechtsprechung⁵⁾ entsteht, so kann man auch verstehen, warum Forderungen, juristische Ar-

1) R.Carnap, Logical Foundations of Probability (1950), S.211f; Induktive Logik und Wahrscheinlichkeit (zusammen mit W.Stegmüller) (1959), S.83.

2) Vgl. Stegmüller, Bd.I, S.660: "Weiß ich um die Richtigkeit von E und verleiht E der Aussage H eine hohe induktive Glaubwürdigkeit, so darf ich daraus noch nicht den Schluß ziehen, daß es vernünftig wäre, H zu akzeptieren. Es kann nämlich der Fall sein, daß ich bei Benützung des für H günstigen Wissens E andere Daten unberücksichtigt gelassen habe, die mir bekannt sind und die gegen H sprechen. In der Prämissen-Terminologie ausgedrückt: Erhielt eine Hypothese auf Grund verfügbaren Wissens einen bestimmten Grad an Glaubwürdigkeit oder Bestätigung, so kann zusätzliches Wissen diesen Bestätigungsgrad verändern. Zahllose Beispiele hierfür, etwa aus der unternehmerischen Planung oder aus dem Gerichtsverfahren, ließen sich anführen. Verfügbares Wissen darf nur dann unberücksichtigt bleiben, wenn es für die Beurteilung der fraglichen Hypothese irrelevant ist"

3) In der oben angeedeuteten juristischen Beurteilungspraxis ist die Sachlage etwas komplizierter: Wir müssen nicht nur das Gesamtdatum, sondern auch die Gesamtglaubwürdigkeit dieses Datums berücksichtigen.

4) Carnaps Forderung nach dem Gesamtdatum bleibt im Prinzip immer theoretisch, solange keine Möglichkeit gegeben ist, die Situation bzw. das Erfahrungswissen sinnvoll einzugrenzen. Bei der obigen Erläuterung (und angewandt auf eine juristische Situation) haben wir die Forderung nach dem relevanten Gesamtdatum aus praktischen Gründen so verstanden, daß das bereits verfügbare Erfahrungswissen nicht vernachlässigt werden darf.

5) Wir bereits auf S.58f dieser Arbeit festgestellt, ist man in der Rechtspraxis primär an der Herstellung der Fiktion "Wahrheit" interessiert. Betrachtet man die gewöhnliche Prozeßpraxis, so zeigt es sich, daß das oben angesprochene Problem der unzureichenden Berücksichtigung des Gesamtdatums - wegen der juristischen Forderung nach der völligen Überzeugung - die übliche juristische Vorgehensweise in der Praxis betrifft: Man tut gewöhnlich so, als ob Informationen nur zugunsten eines Sachverhalts vorliegen. Beweismittel zugunsten eines anderen Sachverhalts werden gewissermaßen "unter den Tisch gekehrt", falls die Schlußfolgerung aufgrund dieser Daten bzw. die Da-

gumentationen wahrscheinlichkeitstheoretisch zu rekonstruieren, in letzter Zeit immer lauter werden.

Diese Problematik wird von denjenigen nicht erkannt, die sich gegen die Anwendung der Wahrscheinlichkeitstheorie auf die Rechtspraxis aussprechen. Gerade in der Vernachlässigung bestimmter Schlußfolgerungen seitens der Rechtsprechung wird eine Rechtfertigung für die Annahme gesehen, daß juristischen Argumentationen entweder ein ganz anders strukturierter oder gar kein Wahrscheinlichkeitsbegriff zugrunde liegt!)

ten selbst weniger glaubwürdig erscheinen. Diese Vorgehensweise, die die Unsicherheit der Beurteilungssituation nicht adäquat berücksichtigt, scheint bereits im Prozeßrecht verankert zu sein; vgl. Berg (S.72f): "Der Richter ist damit durch das Prozeßrecht ermächtigt, bloße (immerhin aber reale) Möglichkeiten und geringe Wahrscheinlichkeiten eines anderen als des für zutreffend gehaltenen Sachverhalts bei der Urteilsfällung beiseite zu lassen. Darin liegt bereits eine Technik zum Abfangen tatsächlicher Ungewißheit". (Hervorhebung im Text). Ergänzen läßt sich, daß nicht nur "bloße Möglichkeiten" und "geringe Wahrscheinlichkeiten" auf diese Weise tatsächlich "abgefangen" werden.

1) Die Auffassung, daß juristische Schlüsse keine Wahrscheinlichkeitsschlüsse sind, unter Berufung darauf, daß ein juristischer Sachverhalt immer im Lichte eines ihn bestätigenden Datums (also unter Vernachlässigung des das Gegenteil stützenden Datums) beurteilt werde bzw. daß der Fall juristisch undenkbar sei, daß ein und dasselbe Datum (das Gesamtdatum tut es fast immer) sowohl den Sachverhalt wie auch sein Gegenteil wahrscheinlich macht, vertreten z.B. Brilmayer&Kornhauser (vgl. Review: Quantitative Methods and Legal Decisions, in: 46 Uni. of Chicago Law Rev. (1978), S.135-148, S.146), ohne die Motivation, die einer solchen Vernachlässigung zugrunde liegt, zu bedenken. (Zu Brilmayer&Kornhauser siehe S.139ff dieser Arbeit).

Cohen (vgl. The Probable and the Provable (1977)) hingegen vertritt die Auffassung, daß juristischen Urteilen wohl ein Wahrscheinlichkeitsbegriff zugrunde liegt; allerdings sei dieser anders strukturiert als der herkömmliche Wahrscheinlichkeitsbegriff. Auch Cohen geht davon aus, daß ein und dasselbe Datum nicht gleichzeitig die Hypothese und ihre Negation wahrscheinlich machen kann. Es ist also nicht möglich, daß die Hypothese (im obigen Beispiel) "Y ist römisch-katolisch" aufgrund des Gesamtdatums "Y ist ein Schwede und Y pilgerte im vorigen Jahr nach Lourdes" und ebenso die Negation der Hypothese, d.h. "Y ist nicht römisch-katolisch" aufgrund desselben Datums wahrscheinlich sind. Stattdessen heißt es bei Cohen (im Rahmen seiner "induktiven Wahrscheinlichkeitstheorie", auf die hier leider nicht näher eingegangen werden kann), daß nur eine dieser Hypothesen (welche?) aufgrund des Gesamtdatums eine positive Wahrscheinlichkeit enthält, die andere hingegen die Wahrscheinlichkeit 0 haben muß. (Das folgt aus dem Cohenschen Negationsprinzip, das besagt: Wenn $P_1(H/D_1 \wedge D_2) > 0$ und $P_1(D_1 \wedge D_2) \neq 0$, dann muß $P_1(\neg H/D_1 \wedge D_2) = 0$ sein; P_1 steht für die Cohensche "induktive Wahrscheinlichkeit", H ist die Hypothese (im obigen Beispiel also "Y ist römisch-katolisch") und $D_1 \wedge D_2$ das gesamte vorliegende Datum (im Beispiel: "Y ist ein Schwede" (D_1) und "Y pilgerte im vorigen Jahr nach Lourdes" (D_2)).

Dieser intuitiv schwer nachvollziehbare Schluß hat vielleicht seine Entsprechung in der oben angesprochenen, juristisch verankerten Praxis, so zu tun, als sprächen alle Daten nur zugunsten einer Hypothese.

Übertragen auf die Jurisprudenz muß also die Forderung des Gesamtdatums bei der Beurteilung einer juristischen Sachverhaltshypothese gelten. D.h. sowohl der Sachverhalt wie auch seine Negation müssen in einem konkreten juristischen Fall aufgrund aller (bereits) zur Verfügung stehenden relevanten Informationen beurteilt werden.

Aufgrund des Gesamtdatums wird festgestellt, ob die Hypothese oder ihre Negation überwiegt. Das bedeutet für die eingangs (S.70) besprochene Situation, daß auch dann, wenn die empirischen Befunde, welche gegen den Sachverhalt sprechen, nicht ganz glaubwürdig sind¹⁾, sie der Beurteilung des Sachverhalts - wegen der Forderung des gesamten relevanten Datums - zugrunde gelegt werden müssen und daß wir daher nie den Überzeugungsgrad gleich 1 ($\hat{=}$ "völlige Überzeugung") vom Bestehen des Sachverhalts erhalten können.

So gesehen, spricht die Tatsache, daß es fast nie bzw. sehr selten der Fall ist, daß alle im Prozeß vorliegenden Beweismittel nur eine Hypothese stützen - in der Regel werden Informationen vorliegen, von denen einige die Sachverhaltshypothese und die anderen ihre Negation bestätigen - gegen die Annahme der Gewißheit und für die der Glaubwürdigkeit als rationalen juristischen Überzeugungsmaßstab.

Aus wahrscheinlichkeitstheoretischer Sicht kann man die "völlige Überzeugung" des Richters nicht rechtfertigen. Aber auch juristisch gesehen fehlt es an einer Rechtfertigungsgrundlage. Denn als das einzige Beurteilungskriterium wird hier das "richterliche Rechtsgefühl" angeführt²⁾ Gefühle sind aber - auch wenn sie sich manchmal als richtig erweisen - irrational. Das Rechtsgefühl als einziges Kriterium kann also keine Rationalität der strafrechtlichen Entscheidungsbildung garantieren. Was bleibt, ist eine ziemlich unüberschaubare, unkontrollierbare und im Grunde willkürliche Entscheidungspraxis, die jedoch trotzdem mit einem weitreichenden Geltungsanspruch auftritt.

1) deren Glaubwürdigkeit jedoch nicht gleich Null ist.

2) So bezeichnet z.B. Engisch das "Rechtsgefühl" als "Maßstab für die Richtigkeit bzw. Unrichtigkeit des Ergebnisses" (S.10) und die "subjektive Evidenz" als "Maßstab für Unterscheidung von Scheinbegründung und echter Begründung" (S.11). K.Engisch, Wahrheit und Richtigkeit im juristischen Denken (1963). Vgl. ebenda S.22: "Der Jurist sieht auf der einen Seite das Feld des Wahren und unzweifelbar Richtigen, auf der anderen Seite das Feld des Unwahren und sicher Fehlsamen".

2.6. Da Juristen in allen (auch offenkündig unsicheren) Rechtsfällen davon sprechen, "sie seien überzeugt" usw., könnte man fragen, ob vielleicht doch ein anderer als der gewöhnliche Wahrscheinlichkeitsbegriff dieser Überzeugung zugrunde liegt. Es hätte dann vielleicht Sinn, einen "juristischen Wahrscheinlichkeitsbegriff" anzunehmen, der nichts mit dem gewöhnlichen (gleichgültig, wie man diesen interpretiert) zu tun hat. In der Tat wurde ein Versuch von L.J.Cohen¹⁾ unternommen, einen derartigen "juristischen Wahrscheinlichkeitsbegriff" zu explizieren.

Da einerseits in dieser Arbeit die Auffassung vertreten wird, daß der subjektive Wahrscheinlichkeitsbegriff die Präzisierung des vorwissenschaftlichen juristischen Wahrscheinlichkeitsbegriffs darstellt, andererseits im modernen juristischen Denken seit der Arbeit von Cohen die Tendenz spürbar ist, einen besonderen juristischen Wahrscheinlichkeitsbegriff, der mit dem mathematischen nichts zu tun hat, zu behaupten, so ist es hier angebracht, dort anzusetzen, wo die Diskrepanzen eigentlich beginnen, nämlich bei der Intuition.

Während wir bei der Explikation des intuitiven juristischen Wahrscheinlichkeitsbegriffs an den alltäglichen Gebrauch anknüpfen, geht Cohen hingegen davon aus, daß es eine "juristische Intuition" gibt.

Man kann sich fragen, ob es einen Sinn hat, von einer besonderen juristischen Intuition zu reden, denn auch eine solche muß allen - und nicht nur Juristen²⁾ - zugänglich sein³⁾ Vielmehr drängt sich

1) L.J.Cohen, The Probable and the Provable (1977).

2) Die meisten Juristen, die die Auffassung vertreten, daß es sich bei juristischen Urteilen um Wahrscheinlichkeitsurteile handelt, orientieren sich allerdings an dem gewöhnlichen, d.h. mathematischen Wahrscheinlichkeitsbegriff. Die Diskrepanzen, die hier auftreten, betreffen lediglich die Interpretation eines solchen juristischen Wahrscheinlichkeitsbegriffs.

3) Cohen beansprucht allerdings, daß diese juristische Intuition eine viel allgemeinere Rolle, etwa auch im Alltag, spielt; vgl.S.3: "I shall try to show that this particular non-Pascalien concept of probability (gemeint ist seine "induktive Wahrscheinlichkeitstheorie" - Anm. von mir) can indeed perform the forensic tasks for which any Pascalien concept (Cohen meint damit die klassische Wahrscheinlichkeitstheorie - Anm. von mir) seems ill-suited. And I shall also try to show that the non-Pascalien concept has some important uses even outside the somewhat stylized institutional framework of legal proceedings. For to show the former without the latter would not sufficiently explain

der Verdacht auf, daß das, was hier mit "juristischer Intuition" umschrieben wird und was Cohen darüber hinaus formaliter zu rechtfertigen versucht, nichts anderes ist als die früher angesprochene "Technik zum Abfangen tatsächlicher Ungewißheit"¹⁾ zum Zweck der Herstellung einer Quasi-Wahrheit.

Wir werden uns mit Cohen später (2.Kap.,§3) noch einmal auseinandersetzen. Es geht dabei um die von Cohen gegen die Wahrscheinlichkeitstheorie aufgestellten Paradoxien und ihre Lösungen.

2.7. Oben haben wir festgestellt, daß, wenn uns Gewißheit fehlt, wir jedoch handeln oder entscheiden müssen, wir intuitiv eine Bewertung von Ereignissen mit Wahrscheinlichkeiten vornehmen. Ferner sprachen wir davon, daß der subjektive Wahrscheinlichkeitsbegriff ein Grundbegriff der rationalen Entscheidungstheorie ist, die einen allgemeinen Rahmen darstellt, innerhalb dessen man von konsistenten Glaubensannahmen bzgl. unsicherer Ereignisse und von sonstigen konsistenten Bewertungen zu rationalen Entscheidungen gelangt.

why people are normally content to have such serious issues determined as they are in the courts".

Wie die Analyse des Beispiels auf S.72 Fn.1 gezeigt hat, ist es jedoch sehr schwer bzw. unmöglich, eine solche Intuition im Alltag nachzuvollziehen. Versuche, derartige (intuitiv unplausible und probabilistisch nicht korrekte) Schlüsse nicht nur in der Rechtspraxis zu behaupten, sondern auch auf den Alltag zu erweitern, erwecken eher die Vermutung, daß hier auf Kosten der Korrektheit und Fairneß anderen Motiven der Vorzug gegeben wird, welche die Vernachlässigung bestimmter ('unpassender') Daten, Schlußfolgerungen und Beweise "rechtfertigen".

Vgl. auch (Cohen, S.44): "Hence, if the mathematicist theory does not fit the mode of reasoning that is proper for lay juries, and jurymen are supposed to carry the principles of their everyday reasonings into the jury-room, the mathematicist theory must be wrong for much of everyday discourse also"; und weiter (S.50): "That is, it is worth while studying what kind of structure a concept of probability must have if it is to serve the relatively specialized and determinate purposes of advocates, jurymen, and judges. For there is good reason to believe that just such a concept must also be in use outside the courts for a vast variety of everyday purposes. Otherwise the whole system of lay justice would be founded on an illusion".

1) Vgl. auch S.71f Fn.5 dieser Arbeit.

Entscheidungen müssen nicht unbedingt auf einem Wahrscheinlichkeitsurteil basieren. Ebenso gut kann ihnen eine Bewertung nicht probabilistischer Art zugrunde liegen, die sozusagen "das Tatsachengewicht mißt" und die wir am besten im Sinn von Keynes als "weight of evidence" bzw. "amount of evidence" bezeichnen¹⁾ Man kann fragen, ob juristischen Entscheidungen anstatt eines Wahrscheinlichkeitsbegriffs nicht doch ein solcher des "Tatsachengewichts" zugrunde liegt. Das wollen wir nun an einem Beispiel näher untersuchen.

Angenommen, gegen den Angeklagten sprechen in einem Rechtsfall ein Motiv und mehrere Zeugenaussagen; für den Angeklagten hingegen nichts. Das Tatsachengewicht ist hier eindeutig gegen den Angeklagten. Der Schluß von den Tatsachen auf die Hypothese, daß der Angeklagte schuldig ist, scheint hier "gerechtfertigt" zu sein.

Nehmen wir aber an, für den Angeklagten spräche ein Alibi. Die Situation ist hier nicht mehr so eindeutig wie vorher. Denn um beurteilen zu können, ob der Angeklagte nun schuldig ist oder nicht, müssen wir offensichtlich wissen, was für uns in diesem konkreten Fall von größerer Bedeutung ist: "Motiv und mehrere Zeugenaussagen" oder "Alibi". Was wir hier also neben dem "Tatsachengewicht" brauchen, ist so etwas wie "Relevanzgrad" einzelner Tatsachen.

Außerdem haben wir bei der obigen Rekonstruktion weder die Glaubwürdigkeit bzw. Zuverlässigkeit der Informationen, die für die vorliegenden Beweismittel relevant sind, berücksichtigt, noch, daß es sich bei "Motiv", "Alibi", "Zeugenaussage" und entsprechend bei den aufgrund dieser Beweismittel zu beurteilenden Sachverhalten ihrerseits um unsichere Ereignisse handelt und folglich, daß die Entscheidung unter Unsicherheit (Risiko) erfolgt. (Denn weil die der Entscheidung zugrundeliegenden Sachverhalte unsicher sind, sind auch Fehlentscheidungen, d.h. "Verurteilung eines Unschuldigen" bzw. "Freispruch eines Schuldigen", möglich).

1) J.M.Keynes, A Treatise on Probability (¹1921, 1957).

Auch Cohen spricht im Zusammenhang mit seiner "induktiven Wahrscheinlichkeitstheorie" von "weight of evidence". Im Unterschied zu Keynes bzw. der hier vertretenen Auffassung versteht Cohen darunter allerdings einen Wahrscheinlichkeitsbegriff.

Auch wenn man die Unsicherheit im obigen Fall verneinen würde, was eigentlich auf die bereits kritisierte "Herstellung der Fiktion der Wahrheit" hinausliefe, bleibt das Problem der "Metrisierung des Relevanzgrades" einzelner Tatsachen weiter bestehen¹⁾.

Und schließlich sei gesagt, daß der Schluß vom Vorliegen der Tatsachen aufgrund des "Tatsachengewichts" auf das Vorliegen der Hypothese (der kein probabilistischer Schluß ist) nur scheinbar gerechtfertigt ist. In Wirklichkeit ist ein solcher Schluß nicht korrekt. Denn das Vorliegen von Daten, welche die Hypothese stützen, impliziert die Hypothese (die in diesen Daten nicht enthalten ist) keineswegs.

1) Theoretisch könnte man die Unsicherheit berücksichtigen, indem man den 'Relevanzgrad' jeder einzelner Tatsache mit der Wahrscheinlichkeit für ihr Vorliegen gewichtet, um auf diese Weise das "gesamte Tatsachen-Relevanz-Gewicht" zu erhalten.

Zusammenfassung und Ausblick

Die Wahrscheinlichkeitstheorie in der subjektivistischen Interpretation liefert uns zusammen mit der rationalen Entscheidungstheorie ein rationales Modell dessen, wie man im Alltag von subjektiven Bewertungen von unsicheren Sachverhalten zu konsistenten Entscheidungen gelangt.

Ist man der Auffassung, daß der juristische Wahrscheinlichkeitsbegriff nichts anderes ist als der vorwissenschaftliche Begriff der Wahrscheinlichkeit, den wir im Alltag und sonst beim induktiven (intuitiven) Rasonieren verwenden und dessen Präzisierung der subjektive Wahrscheinlichkeitsbegriff darstellt, dann ist auch die rationale, d.h. entscheidungstheoretische Rekonstruktion der Rechtspraxis als einer solchen unter Unsicherheit (Risiko) möglich.

Die Probleme entstehen erst dann, wenn bei der Rekonstruktion der Rechtspraxis von einem Wahrscheinlichkeitsbegriff ausgegangen wird, der eine ganz andere Struktur hat als der gewöhnliche. Man muß dann nämlich erstens diesen ohne Bezug auf den alltäglichen intuitiven Wahrscheinlichkeitsbegriff rechtfertigen und zweitens die Rationalität der Entscheidungen, denen dieser Wahrscheinlichkeitsbegriff zugrunde liegt, begründen. Beides kann aber nicht gelingen. Ersteres scheitert an der Unglaubwürdigkeit der Annahme einer besonderen Intuition, die lediglich Juristen zu eigen ist, darüber hinaus auch an dem erkenntnistheoretischen Esoterismus, den eine solche Annahme impliziert; das zweite an der Unmöglichkeit, die Rationalität von Entscheidungen angesichts der Unsicherheit anders als in einem entscheidungstheoretischen Rahmen, d.h. einer Theorie deren Grundbegriff der subjektive Wahrscheinlichkeitsbegriff ist, zu begründen.

Anknüpfend an diese theoretischen Erörterungen wollen wir im nächsten Kapitel auf die Anwendung der Wahrscheinlichkeitstheorie auf die Strafrechtspraxis näher eingehen. Wir beginnen mit der Schilderung der Diskussion, die in den letzten 25 Jahren um den juristischen Wahrscheinlichkeitsbegriff in der juristischen (vorwiegend anglo-amerikanischen) Literatur stattgefunden hat.

2. Kapitel: Anwendung der Wahrscheinlichkeitstheorie auf die Strafrechtspraxis

Einleitung

Seit ca. 25 Jahren ist ein ständig steigendes Interesse praktisch arbeitender Juristen an wahrscheinlichkeitstheoretischen, insbesondere statistischen Methoden zu beobachten. Die Beschäftigung damit begann im anglo-amerikanischen Sprachraum¹⁾ aber auch in Skandinavien²⁾ Mit den Arbeiten von Kegel³⁾ und Eggleston⁴⁾ traten auch Deutschland bzw. Australien in die Diskussion des Wahrscheinlichkeitsbegriffs in der Jurisprudenz ein.

Die Diskussion des Wahrscheinlichkeitsbegriffs in der Rechtspraxis bzw. seiner Brauchbarkeit für juristische Zwecke geht natürlich weiter zurück; vgl. J.Bentham, A Treatise on Judicial Evidence (1825), S.D.Poisson, Recherches sur la probabilité des jugements, etc. (1837), G.Boole, On the Application of the Theory of Probabilities to the Question of the Combination of Testemonies or Judgments (1857), J.H.Wigmore, Evidence (1913); im deutschen Sprachraum: J.v.Kries, Über

1) V.C.Ball, The Moment of Truth: Probability Theory and Standards of Proof (1961) E.B.Mode, Probability and Criminalistics (1963); C.R.Kingston und P.L.Kirk, The Use of Statistics in Criminalistics (1964); C.R.Kingston, Applications of Probability Theory in Criminalistics (1965), Probability and Legal Proceedings (1966) J.B.Parker, A Statistical Treatment of Identification Problems (1966); W.B.Stoe-buck, Relevancy and the Theory of Probability (1966); T.H.Liddle, Mathematical and Statistical Probability As a Test of Circumstantial Evidence (1968); P.Sigal, Judicial Use, Misuse, and Abuse of Statistical Evidence (1969); A.D.Cullison, Identification by Probabilities and Trial by Arithmetic (A Lesson for Beginners in How to be Wrong with Greater Precision) (1969); M.O.Finkelstein&W.B.Fairley, A Bayesian Approach to Identification Evidence (1970); L.H.Tribe, Trial by Mathematics: Precision and Ritual in the Legal Process (1971); R.J.Simon&L.Mahan, Quantifying Burdens of Proof (1971); R.F.Coleman&H.J.Walls, The Evaluation of Scientific Evidence (1974); R.L.Smith&R.P.Charrow, Upper and Lower Bounds for Probability of Guilt Based on Circumstantial Evidence (1975); D.V.Lindley, Probabilities and the Law (1975); R.O.Lempert, Modeling Relevance (1977); E.Gerjuoy, The Relevance of Probability Theory to Problems of Relevance (1977); L.Brilmayer&L.Kornhauser, Review: Quantitative Methods and Legal Decisions (1978); D.Kaye, The Laws of Probability and the Law of the Land (1979), Probability Theory Meets Res Ipsa Liquitur (1979); D.A.Schum, A Review of a Case Against Blaise Pascal and his Heirs (1979); J.D.Jackson, Probability and Mathematics in Court Fact-finding (1980); D.Stripinis, Probability Theory and Circumstantial Evidence: Implications from a Mathematical Analysis (1981).
(Die oben angeführte Liste der Aufsätze hat keinen Anspruch auf Vollständigkeit).

2) P.O.Ekelöf, Free Evaluation of Evidence (1964).

3) Vgl. S.65 Fn.1, die deutschen Beiträge wurden bereits auf S.63ff dieser Arbeit kurz angesprochen; dort findet der Leser auch die genauen Titel der Beiträge.

4) R.Eggleston, Probabilities and Proof, in: 4 Melbourne Univ. Law Rev. (1963); Evidence, Proof and Probability (1978), Beyond reasonable doubt (1977), Subjective Probability and the Law (1979), The Probability Debate (1980).
("Evidence, Proof and Probability" ist eines der drei Bücher, die überhaupt zur Anwendung der Wahrscheinlichkeitstheorie in der Rechtspraxis erschienen sind; die zwei anderen sind von L.J.Cohen, The Probable and the Provable (1977) und M.O.Finkelstein, Quantitative Methods in Law (1978)).

die Begriffe der Wahrscheinlichkeit und Möglichkeit und ihre Bedeutung im Strafrecht (1889). Siehe auch G.F.James, *Relevancy, Probability and the Law* (1941); J.P.McBaine, *Burden of Proof: Degrees of Belief* (1944).

Wigmore weist als erster darauf hin, daß die Form der vor Gericht gebrachten Argumentationen induktiv ist und daß juristische Urteile - ebenso wie die alltäglichen - Wahrscheinlichkeitsurteile sind. James greift die Ideen von Wigmore auf, indem er untersucht, wie neue Beweismittel die ursprüngliche Wahrscheinlichkeit für die Schuld des Angeklagten verändern können; in diesem Zusammenhang nimmt er auch Stellung zum Begriff der Relevanz als Veränderung der ursprünglichen Wahrscheinlichkeit. James vertritt ebenso wie Wigmore die Auffassung, daß es keine strikte Ableitung im Sinn der logischen Subsumtion für juristische Beweise gibt. McBaine ist der Meinung, daß juristische Beweise als Glaubensurteile Wahrscheinlichkeitsurteile sind. Er lehnt die Auffassung ab, wonach Tatsachen mit Sicherheit nachgewiesen werden können; vgl. (S.250): "...that fact finding at most is only a process of determining what probably, in varies degrees, has happened".

Wir wollen auf die Diskussion dieser Literatur nicht weiter eingehen, weil dort lediglich informelle Betrachtungen zum Wahrscheinlichkeitsbegriff angestellt werden, dagegen keine systematische Stellungnahme zur Natur des Begriffs der Wahrscheinlichkeit und zu seiner Deutung stattfindet.

Dagegen scheint der Aufsatz von Ball¹⁾ als Einstieg für unsere Diskussion (§1) geeignet zu sein. Denn bei ihm finden wir nicht nur bereits alle Ansätze einer modernen Auffassung bzgl. der Anwendung der Wahrscheinlichkeitstheorie auf die Rechtspraxis, sondern zum ersten Mal auch den Gebrauch einiger Techniken aus der Wahrscheinlichkeitsrechnung²⁾

Während man im anglo-amerikanischen Raum, in Skandinavien und Australien primär am technischen Aspekt der Wahrscheinlichkeitstheorie interessiert ist - man sucht nach probabilistischen und statistischen Verfahren und Techniken, mit deren Hilfe die Informationsverarbeitung bei der Beweiswürdigung im Prozeß präzisiert und rationalisiert werden kann - trägt die Diskussion des Wahrscheinlichkeitsbegriffs in Deutschland eher einen kritisch-polemischen Charakter - hier will man primär den Sinn und die Existenzberechtigung der Verwendung von Wahrscheinlichkeitsurteilen in der Rechtspraxis klären³⁾. Es ist daher nicht verwunderlich, daß man im deutschen Sprachraum viel skeptischer und ablehnender der Benutzung statistischer bzw. wahrscheinlichkeitstheoretischer Methoden bei der Beweiswürdigung, oder allgemeiner: der Anwendung der Wahrscheinlichkeitstheorie auf die Rechtspraxis, gegenüber steht.

1) Vgl. S. 79 Fn.1 dieser Arbeit.

2) Vgl. aber auch G.Boole, *On the Application of the Theory of Probabilities to the Question of the Combination of Testemonies or Judgements* (1857); I.J.Good, *Probability and the Weighing of Evidence* (1950).

3) Einen mehr oder weniger technischen Charakter trägt lediglich der Aufsatz von Schreiber, vgl. S. 103ff dieser Arbeit. Mit der Anwendung der Wahrscheinlichkeitstheorie auf die Rechtspraxis befaßt sich, allerdings auf eine informelle Weise, Kegel, vgl. S. 98ff dieser Arbeit.

Auf die Anwendung der Bayes-Technik gehen im deutschen Sprachraum ein: Bender/Röder/Nack, S. 191ff (vgl. S. 66 dieser Arbeit), Koch/Rüßmann, *Juristische Begründungslehre* (1982), S. 318ff, E.L.Nell (vgl. S. 62 dieser Arbeit), S. 50ff und R.Motsch, *Vom rechtsgenügenden Beweis* (1983), S. 157ff. Zur Anwendung der Bayes-Methode in der (deutschen) kriminologischen Forschung vgl. den Aufsatz von K.D.Nissen, "Theorie der Begutachtung" (1983), S. 143-152, in: 172 *Archiv für Kriminologie*, 1983.

Vieles spricht dafür, daß die oben angesprochene Entwicklung noch im Anfangsstadium steckt und daß die juristische Diskussion des Wahrscheinlichkeitsbegriffs noch längst nicht abgeschlossen ist. Das dies so ist, ist etwa daraus zu ersehen, daß man sich noch immer um die Klärung des Zusammenhangs von statistischen und nicht statistischen Daten bemüht und daß man noch immer nach passenden statistischen Methoden sucht, mittels derer man die in der Rechtspraxis oft auftretenden Identifizierungsprobleme lösen will. Daß dabei auch statistisch versierte Spezialisten vor Fehlern nicht geübt sind, veranschaulicht die 10 Jahre lang währende Diskussion des vielleicht bekanntesten Falls in der anglo-amerikanischen Rechtsliteratur, nämlich des Falls 'People v. Collins'¹⁾

Daß diese Entwicklung sich noch im Anfangsstadium befindet, zeigt auch die Tatsache, daß man sich zwar mit statistischen Methoden auseinandersetzt, sich über die Natur des juristischen Beweises selbstgemeint ist der Zusammenhang zwischen Daten und der jeweiligen Hypothese über die Täterschaft – wohl kaum Gedanken macht bzw. davor zurückschreckt, diesen Zusammenhang wahrscheinlichkeits-theoretisch zu behandeln²⁾

Und nicht zuletzt weist auf den Anfangscharakter dieser Entwicklung auch die Tatsache hin, daß, während es an Aufsätzen zum Wahrscheinlichkeitsbegriff in der Rechtsliteratur nicht mangelt, die naheliegende Verbindung zur Entscheidungstheorie übersehen bzw. nicht in gebührender Weise berücksichtigt wird. Daß gerade die Rechtspraxis wie sonst keine andere praktisch operierende Wissenschaft ein Musterbeispiel einer entscheidungstheoretisch orientierten und rational operierenden Wissenschaft abgeben könnte, wird anscheinend nicht gesehen. Abgesehen von ein paar unzureichenden Modellen³⁾ gibt es kein einigermaßen korrekt funktionierendes, adäquates Entscheidungsmodell⁴⁾, geschweige denn eine rationale juristische Entscheidungstheorie.

1) vgl. S.112ff dieser Arbeit.

2) Vgl. Finkelstein&Fairley; S.123ff, die das Bayes-Theorem in die Rechtspraxis erstmals explizit eingeführt haben, allerdings mit der Einschränkung, daß dieses nur bei der Feststellung von empirischen Befunden bzw. Beweismitteln anzuwenden sei.

3) Siehe J.Kaplan, Decision Theory and the Factfinding Process (1968); B.M.Maassen, Beweisprobleme im Schadensersatzprozeß (1975), S.7f; R.Motsch, Vom Prozeß als Beweis zum Überwiegensprinzip (1978); C.R.Marshall&J.A.Wise, Jurors Decisions and the Determination of Guilt in Capital Punishment Cases: A Bayesian Perspective (1975); A.D.Cullison, The Model of Rules and the Logic of Decision (1977); S.S.Nagel&M.G.Neef, Decision Theory and the Legal Process (1979), Kap.1. (Auf diese Modelle wird im 2.Teil dieser Arbeit noch näher eingegangen). Siehe auch: R.O.Lempert, Modeling Relevance, in: 75 Michigan Law Rev. (1977), S.1021-1057.

4) Die Modelle von Marshall&Wise und Nagel&Neef sind zwar konsistent, geben jedoch die juristische Entscheidungspraxis nicht adäquat wieder, da sie weder die Tatsache berücksichtigen, daß es mehr als 2 Tatbestände und dementsprechend mehr als 2 Entscheidungsmöglichkeiten geben kann, noch die Möglichkeit ausschließen, daß die Verurteilung gewählt wird, wenn die Wahrscheinlichkeit für die Täterschaft kleiner ist als eine solche für die Nichttäterschaft. Zum Modell von Marshall&Wise siehe S.157ff und den II.Teil dieser Arbeit; zu Nagel&Neef, S.265 f dieser Arbeit.

Gerade die von mir zuletzt angesprochenen Gesichtspunkte - die wahrscheinlichkeitstheoretische Beurteilung des Sachverhalts aufgrund von Daten und die rationale Entscheidungsbildung, der die solchermaßen festgestellten Sachverhalte zugrunde liegen - zeigen die Notwendigkeit, sich mit diesen Themen auseinanderzusetzen. Es soll daher in diesem Kapitel der Versuch unternommen werden, juristische Beweise wahrscheinlichkeitstheoretisch zu rekonstruieren (§1). Im 2. Teil dieser Arbeit wird darüber hinaus eine axiomatisch aufgebaute rationale juristische Entscheidungstheorie entwickelt, die vom wahrscheinlichkeitstheoretisch festgestellten Sachverhalt ausgeht und zur rationale Wahl einer juristischen Maßnahme führt. Außerdem wollen wir eine numerische Berechnung der tatschuldangemessenen Strafe innerhalb des Strafrahmens vornehmen. Zuerst wird jedoch ein Bild dessen vermittelt, was zur Anwendung der Wahrscheinlichkeitstheorie auf die Rechtspraxis in der juristischen Literatur bis jetzt geäußert wurde.

Ich werde somit zuerst die Diskussion des Wahrscheinlichkeitsbegriffs schildern, wobei durch die Auseinandersetzung mit verschiedenen Meinungen auch die eigene Position weiter geklärt und vertieft werden soll (§1). Wie schon aus dem vorangegangenen Kapitel klar geworden sein dürfte, vertrete ich die Auffassung, daß es sich bei dem juristischen Beweisen zugrundeliegenden Wahrscheinlichkeitsbegriff um einen subjektiven handelt. Dieser Wahrscheinlichkeitsbegriff ist die Präzisierung des intuitiven Begriffs der Überzeugung, der juristischen Entscheidungen zugrunde liegt und der durch die juristische, auf die Herstellung der Fiktion "Wahrheit" gerichtete, Vorgehensweise nur verdeckt wird¹⁾. Außerdem bin ich der Meinung, daß der statistische Wahrscheinlichkeitsbegriff zwar eine Rolle in der Rechtspraxis spielen kann, allerdings nur bei der Beurteilung statistischer Daten, nicht dagegen bei einer solchen der Sachverhaltshypothese; v.a. soll er auf keinen Fall "die Überzeugungsbildung ersetzen"²⁾.

Die Auffassung, daß es sich bei der juristischen Überzeugungsbildung um eine solche in subjektiven Wahrscheinlichkeitsurteilen handelt, wird durch eine Reihe empirischer Untersuchungen bestätigt. In §2 werden wir auf vier Arbeiten, die sich mit dieser Problematik befassen, näher eingehen. Anhand dieser Arbeiten wird gezeigt, daß so manche übliche Vorstellung über die Urteilsbildung praktisch arbeitender Juristen revidiert werden muß, wie etwa, daß Laien- und Berufsrichter mit Wahrscheinlichkeiten nicht umgehen können, daß die intuitive Verarbeitung von Informationen zum Gesamtbild in einem Rechtsfall dem Bayesschen Ansatz widerspricht³⁾ und daß die "an Sicherheit grenzende Wahrscheinlichkeit" in den Augen derjenigen, die juristische Fälle beurteilen, entweder als Wahrscheinlichkeit nicht präzisierbar ist, weil darunter "nur erkenntnistheore-

1) Zu diesem Punkt siehe S. 58f, 70ff dieser Arbeit.

2) Entgegen der Auffassung von Schneider bzw. Berg, daß bei manchen juristischen Beweisen "die Statistik die Überzeugung ersetzt" (S. 65 dieser Arbeit), wird hier die Meinung vertreten, daß die juristische Überzeugungsbildung letztlich subjektiv ist, auch dann, wenn manche relevanten Daten statistisch ausgewertet werden.

3) Dies ist ungefähr der Standpunkt von Tribe (Siehe S. 128ff dieser Arbeit), der behauptet, daß die Anwendung der Wahrscheinlichkeitstheorie auf die Rechtspraxis alles andere als sinnvoll sei, aber auch der von Cohen, der (im Unterschied zu Tribe) allerdings meint, daß hier prinzipiell andersgeartete Wahrscheinlichkeitsurteile vorlägen.

tische Zweifel" zu verstehen sind, oder, numerisch ausgedrückt, so etwas wie 0.99... (eine weitere in der modernen juristischen Literatur verbreitete quantitative Umschreibung des obigen Begriffs, die ausdrücken soll, daß "nur theoretische Zweifel bestehen") darstellt¹⁾

Einen anderen Ansatz bei der Behandlung juristischer Urteile als Wahrscheinlichkeitsurteile hat, wie bereits erwähnt, Cohen gewählt. Nach Cohen handelt es sich bei dem juristischen Wahrscheinlichkeitsbegriff weder um einen alltäglichen, vorwissenschaftlichen, der durch den Begriff der subjektiven Wahrscheinlichkeit präzisiert werden kann, noch um einen objektiven (statistischen) Wahrscheinlichkeitsbegriff. Der juristische Wahrscheinlichkeitsbegriff sei ganz anders strukturiert. In seinem Buch versucht Cohen die Intuition formal zu präzisieren, die diesem "juristischen Wahrscheinlichkeitsbegriff" zugrunde liegen soll. Seine Kritik gründet sich dabei darauf, daß der mathematische Wahrscheinlichkeitsbegriff, angewandt auf die Rechtspraxis, zu Paradoxien führe und daher ungeeignet sei, diese adäquat zu behandeln. Bei näherem Hinsehen erweisen sich jedoch diese Paradoxien als Scheinparadoxien; deren Auflösung entzieht somit der sog. "induktiven" Theorie Cohens gewissermaßen eine starke Rechtfertigungsgrundlage. Ich werde diese Paradoxien mit ihren Lösungen in §3 darstellen.

In §4 werden zwei juristische Fälle - 'People v. Collins' und ein fiktiver Fall - wahrscheinlichkeitstheoretisch rekonstruiert. Vier Formeln erweisen sich dabei als wichtig: das Multiplikationsprinzip, das Bayes-Theorem, die B_{erw} -Formel (eine Erweiterung der Bayes-Formel²⁾) und die Jeffrey-Formel. Außerdem werden die für die Wahrscheinlichkeitstheoretische Rekonstruktion der Rechtspraxis wichtigen Forderungen, wie die des relevanten Datums, die des gesamten Datums usw., angegeben, Begriffe wie "Apriori-Wahrscheinlichkeit" und Zeugenglaubwürdigkeit (Datenzuverlässigkeit) erläutert und der Begriff der (juristischen) Relevanz neu definiert.

1) Vgl. S. 58, Fn.1 dieser Arbeit.

2) Vgl. S. 34 dieser Arbeit.

Sl. Diskussion des Wahrscheinlichkeitsbegriffs in der juristischen Literatur

1.1. Ball¹⁾

Bei Ball finden wir bereits alle Ansätze einer modernen Auffassung bzgl. der Anwendung der Wahrscheinlichkeitstheorie auf die Rechtspraxis. So schreibt er, daß der Beweis- und Überzeugungsprozeß in der Rechtspraxis zum Bereich der Anwendung der Wahrscheinlichkeitstheorie gehört²⁾ und außerdem, daß es eine juristische Konvention ist, Wahrscheinlichkeitsurteile als Wahrheitsbehauptungen zu behandeln. Das einzige, was wir in bezug auf juristische Sachverhalte erreichen können, ist nach Ball, ihnen lediglich Wahrscheinlichkeitsgrade zuzuordnen³⁾ Ball versucht außerdem als erster in Anlehnung an George Boole⁴⁾ juristische Probleme formal zu behandeln. So führt er auf S.819 einen Fall auf, in welchem Zeugenaussagen eine entscheidende Rolle spielen, und den er wahrscheinlichkeitstheoretisch löst⁵⁾

Es handelt sich hier um den Fall einer Fälschung. A verklagt B anhand eines Schecks, den B angeblich unterschrieben hat. B bestreitet das und behauptet, daß jener eine Fälschung sei. Es sind 11 Zeugen (Handschriftenexperten) eingeladen; 6 davon bezeugen, daß die Unterschrift diejenige von B ist, und 5, daß sie das nicht ist.

Als erstes beschreibt Ball ein Experiment von Inbau⁶⁾ zur Ermittlung der Zeugenglaubwürdigkeit. Dies geschieht empirisch wie folgt:

Den Zeugen werden in mehreren Durchgängen verschiedene Unterschriften vorgelegt, aus welchen sie ersehen sollen, ob die Unterschriften gefälscht sind oder nicht. Es werden insgesamt 95 Untersuchungen und 380 Identifizierungen gemacht. Die durchschnittliche Treffsicherheit wird mit 63% errechnet.

Diese Zahl legt Ball auch seiner Zeugenglaubwürdigkeit zugrunde. Alle Zeugen werden somit als "gleich glaubwürdig" bzw. "gleich kompetent" angesehen.

1) V.C.Ball, The Moment of Truth: Probability Theory and Standards of Proof, in: 14 Vanderbilt Law Review (1961), S.807-830. .

2) Vgl. ibid. S.807.

3) Vgl. ibid. S.808: "By the time two contradictory material propositions reach the stage of determination, some required degree of probability is all that can be assigned to them...the probable proposition is converted into "juridical truth", for the purpose of settling the dispute with finality and permitting the other affairs of the litigant and the world to go forward".

4) Vgl. S. 79 dieser Arbeit.

5) Vgl. Ball, S.824f, S.826ff.

6) Inbau, Lay Witness Identification of Handwriting, 34 Ill.Law.Rev. (1939), S.433ff.

Als nächstes ermittelt Ball aufgrund der obigen Information (d.h. daß 6 Zeugen für die Echtheit und 5 dagegen sprechen, und daß alle gleich glaubwürdig sind, wobei diese Glaubwürdigkeit 0.63 beträgt) die Wahrscheinlichkeit dafür, daß B den Scheck unterschrieben hat. Das wird anhand einer von G.Boole¹⁾ vorgeschlagenen Methode erreicht.

Die Wahrscheinlichkeit dafür, daß jeder Zeuge korrekt attestiert hat, sei P, dementsprechend ist 1-P die Wahrscheinlichkeit, daß er einen Fehler begangen hat. Ferner seien alle Identifizierungen unabhängig voneinander. Die Boolesche Formel lautet dann:

$$\frac{P^6(1-P)^5}{P^6(1-P)^5 + P^5(1-P)^6}.$$

Mit $P = 0.63$ erhalten wir das Ergebnis 0.63, d.h. die Wahrscheinlichkeit dafür, daß B den Scheck unterschrieben hat aufgrund der Information, daß 6 Zeugen die Echtheit attestieren und 5 die Falschheit, wobei alle gleichermaßen (zu 63%) glaubwürdig sind, ist 0.63.

Dieses Resultat ist nicht verwunderlich, wenn man berücksichtigt, daß ein (zu 63%) glaubwürdiger Zeuge mehr zugunsten der Echtheit der Unterschrift aussagt (die anderen Zeugenaussagen heben sich - da sie alle gleichermaßen glaubwürdig sind - sozusagen gegenseitig auf).

Die Boolesche Formel erinnert stark an das Bayes-Theorem²⁾, wenn wir die Apriori-Wahrscheinlichkeit für die Echtheit der Unterschrift von B mit 0.5 ansetzen; $P^6 = 0.63^6$ ist dann die Wahrscheinlichkeit, daß 6 (voneinander unabhängige) Zeugenaussagen, die zugunsten der Echtheit attestieren, korrekt sind unter der Voraussetzung, daß die Hypothese (über die Echtheit) stimmt; $(1-P)^5 = 0.37^5$ ist die Wahrscheinlichkeit, daß 5 (voneinander unabhängige) Zeugenaussagen, die gegen die Echtheit sprechen, nicht korrekt sind unter der Voraussetzung, daß die Hypothese stimmt; und entsprechend ist $(1-P)^6 = 0.37^6$ die Wahrscheinlichkeit, daß die (voneinander unabhängigen) Zeugenaussagen, die zugunsten der Echtheit attestieren, nicht korrekt sind, wenn die Hypothese (über die Echtheit) nicht stimmt und $P^5 = 0.37^5$ die Wahrscheinlichkeit, daß die Zeugenaussagen, die ge-

1) G.Boole, Studies in Logic and Probability (1952), S.310 (¹1857).

2) Die Boolesche Formel ist trotzdem nicht korrekt; dazu später auf S.109f dieser Arbeit.

gen die Echtheit sprechen, korrekt sind unter der Voraussetzung, daß die Hypothese nicht stimmt.

Es ist jedoch nicht die Absicht Balls, die obige Formel Bayesianistisch zu deuten. So vermeidet er konsequent das Sprechen über die Apriori-Wahrscheinlichkeit der Hypothese¹⁾.

Wie wir gesehen haben, vertritt Ball in bezug auf die Zeugenglaubwürdigkeit den frequentistischen Standpunkt²⁾, denn diese wird gleich der Treffsicherheit gesetzt (d.h. der relativen Häufigkeit der Antworten, bei denen der Zeuge korrekt attestiert).

Abgesehen von der Unangemessenheit der frequentistischen Interpretation bei der gewöhnlichen Bestimmung der Zeugenglaubwürdigkeit³⁾ ist es sicherlich ein Verdienst von Ball, als erster darauf explizit hingewiesen zu haben, daß bestimmte juristische Argumentationen formal nachvollziehbar sein müssen, wenn man sie exakt beurteilen will.

1) Die Boolesche Formel kann nicht bayesianistisch korrekt interpretiert werden, wenn es sich um mehr als zwei gleichwahrscheinliche Hypothesen handelt, resp. wenn die Apriori-Wahrscheinlichkeit der Hypothese $\neq 0.5$ ist.

2) Auch sonst (also nicht nur in bezug auf die Zeugenglaubwürdigkeit) ist Ball der Meinung, daß Gerichte "eine Art der Häufigkeitstheorie" anwenden; vgl. S.812: "If we examine the courts' handling of problems of relevancy, we find that a form of frequency theory is continually applied to determine whether a particular piece of information changes the probability of a material proposition".

Die in der frequentistischen Theorie auftretende Schwierigkeit bzgl. der Behandlung von Einzelfällen versucht er im Sinne von Reichenbach zu lösen (vgl. S.812); zum Reichenbachschen Lösungsversuch vgl. S.40 dieser Arbeit.

3) Denn der von Ball in Anlehnung an Inbau geschilderte Fall stellt eine seltene Ausnahme dar. In diesem Fall läßt sich die Treffsicherheit eines Handschriftenexperten relativ leicht experimentell (in einer Stichprobe) feststellen und extrapolieren. In der Regel, insbesondere, wenn die Zeugen keine Experten sind, deren Kenntnisse und Fertigkeiten man empirisch überprüfen kann, und es sich um einmalige, nicht nachprüfbare Ereignisse handelt, ist eine solche experimentelle Bestimmung der Zeugenglaubwürdigkeit nicht durchführbar. In diesen Fällen ist nur die konventionelle, subjektive Beurteilung der Zeugenglaubwürdigkeit aufgrund der eigenen Erfahrung und der zur Verfügung stehenden Informationen möglich.

1.2. Ekelöf¹⁾

Ekelöf beschäftigt sich mit der Frage, wie das Gericht zum Wissen über relevante Tatsachen kommt. Er vertritt die Ansicht, daß man keine völlige Zuverlässigkeit bzgl. der relevanten Tatsachen erlangen kann und daß man höchstens von einem hohen Wahrscheinlichkeitsgrad hinsichtlich der Feststellung des Sachverhalts, nie aber von dessen Wahrheit sprechen kann. Wenn das Gericht mit zwei kontroversen Auffassungen konfrontiert wird, so ist es nach Ekelöf oft nicht imstande, zu entscheiden, welche von diesen wahr ist und daher gezwungen, die wahrscheinlichere zu wählen. Auch Ekelöf²⁾ geht auf das Problem ein, das sich ergibt, wenn zusätzliche Informationen zur Beurteilung der Hypothese herangezogen werden wie etwa in dem von ihm geschilderten Beispiel auf S.52, wo das Gericht entscheiden soll, ob der Angeklagte der Vater des Kindes ist.

Es wird in diesem Beispiel vereinfacht angenommen, daß die Mutter des Kindes zur Empfängniszeit mit zwei Männern intimen Kontakt hatte, mit dem Angeklagten und mit einem anderen, über den allerdings überhaupt keine Informationen vorliegen. Weiter sei angenommen, daß über den Angeklagten nur eine einzige Information in Form eines genetischen Tests vorliegt, die besagt, daß das Kind von seinem Vater ein bestimmtes Merkmal erben muß, das nur bei 5% der Bevölkerung anzutreffen ist, und daß der Angeklagte dieses Merkmal besitzt.

Die Frage stellt sich somit, wie wahrscheinlich die Vaterschaft des Angeklagten ist aufgrund dieser (statistischen) Information. Diese wird von Ekelöf anhand der Essen-Möller-Formel³⁾ bestimmt, im obigen Fall:

$$\frac{1}{1+0.05} = 0.9524^4)$$

1) P.O.Ekelöf, Free Evaluation of Evidence, in: Scandinavien Studies in Law (1964), S.45-66.

2) Vgl. Ball, der auf das Problem der Berücksichtigung zusätzlicher Informationen (wie Zeugenaussagen) bei der Beurteilung der Hypothese eingeht; S.84f dieser Arbeit.

3) Vgl. Ekelöf, S.52. Ekelöf gibt nur einen rechnerischen Ausdruck der Essen-Möller-Formel an (vgl. Ekelöf, S.52, Fn.4), ohne den Rechenvorgang bzw. die Formel zu begründen bzw. zu erklären. Eric Essen-Möller, Die Beweiskraft der Ähnlichkeit im Vaterschaftsnachweis (1938), in: Mitteilungen der Anthropologischen Gesellschaft in Wien, Bd.LXVIII, S.9-53.

4) Essen-Möller gibt folgende Formel an:

$$\frac{x}{x+y} \quad \text{bzw., nach Umformung,} \quad \frac{1}{1+\frac{y}{x}}$$

wobei x die relative Häufigkeit des kindlichen Merkmals unter wahren Vätern, y eine solche unter falschen Vätern (Essen-Möller setzt diese der relativen Häufigkeit des kindlichen Merkmals in der Gesamtbevölkerung gleich) ist. $\frac{x}{x+y}$ ist somit die Wahrscheinlichkeit für die wahre Vaterschaft unter allen Vätern, die das gleiche Merkmal aufweisen wie das Kind. (vgl.Essen-Möller, S.11ff).

Daraus folgert Ekelöf, daß die Wahrscheinlichkeit dafür, daß der Angeklagte der Vater ist, über 95% beträgt¹⁾

Das obige Verfahren scheint in einem Fall, wo bekannt wird, daß auch der zweite Verdächtige das gleiche Merkmal aufweist, nicht mehr zu funktionieren. Denn in einem solchen Fall ergibt die Essen-Möller-Formel auch für die Vaterschaft des zweiten Verdächtigen die Wahrscheinlichkeit 0.95. Da in diesem Fall das Vorhandensein des gemeinsamen Merkmals völlig irrelevant für die Beurteilung der Vaterschaftsfrage ist, dürfte sich eigentlich die Aposteriori-Wahrscheinlichkeit nicht ändern²⁾ Das Essen-Möller-Verfahren, das von der statistischen Verteilung eines Merkmals in einer Population ausgeht, kann das anscheinend nicht zustandebringen³⁾

1) Vgl. Ekelöf, S. 52: "We assume that this admits the conclusion that there is a probability amounting to 95 per cent for the paternity of the defendant".

2) Man kann überhaupt fragen, inwiefern dann in einem solchen Fall die statistische Beurteilung angemessen ist. Stellt sich nämlich heraus, daß der zweite in Frage kommende Verdächtige nicht dieses Merkmal aufweist, dann folgt aus der Voraussetzung, daß nur diese zwei Männer in Frage kommen, daß das Kind ein bestimmtes Merkmal vom Vater haben muß und ein solches Merkmal sowohl beim Kind wie auch bei dem ersten Verdächtigen gefunden wurde, die Wahrscheinlichkeit gleich 1 dafür, daß dieser der Vater ist.

3) Das Essen-Möller-Verfahren scheint nur dann zu funktionieren, wenn wir unter mehreren Verdächtigen einen einzigen Verdächtigen mit einem bestimmten, in Frage kommenden Merkmal haben und sonst keine Informationen über die anderen Verdächtigen vorliegen. Dieses Verfahren berücksichtigt außerdem in gar keiner Weise die Apriori-Information über die Anzahl der potentiellen Väter (ob 10 oder nur ein einziger verdächtigt wird, spielt hier keine Rolle) und auch nicht andere Informationen über den Verdächtigen.

Die Essen-Möller-Formel ist eine Art statistisch interpretierte Bayes-Formel (mit der Apriori-Wahrscheinlichkeit gleich 0.5). Die oben erwähnte Schwierigkeit betrifft allerdings nicht die Bayes-Formel selbst, sondern vielmehr die Interpretation dieser Formel in statistischen Wahrscheinlichkeiten.

Die subjektivistische Interpretation des oben geschilderten Falls macht dagegen offensichtlich keine Schwierigkeiten: Denn angesichts der Information, daß der zweite Verdächtige das gleiche Merkmal besitzt und daß nur zwei Männer als Verdächtige in Frage kommen, folgt, daß auch dann, wenn der erste Verdächtige nicht der Vater ist, das Kind mit praktischer Sicherheit das fragliche Merkmal wie dieser Verdächtige haben wird (dasselbe gilt für den zweiten Verdächtigen). Die Wahrscheinlichkeit für die Vaterschaft des ersten bzw. des zweiten Verdächtigen ist aufgrund der Information über das kindliche Merkmal daher gleich 0.5, d.h. diese Information erweist sich in diesem konkreten Fall für den Vaterschaftsnachweis als irrelevant.

Schauen wir uns jetzt die Regel an, die Ekelöf zur Ermittlung der Wahrscheinlichkeit für das gemeinsame Vorliegen von zwei Beweismitteln vorschlägt. Auf S.58 gibt er folgendes Beispiel an:

In einen Autounfall sind zwei Autos verwickelt. Es gibt lediglich zwei Informationen, die beweisen sollen, daß eines der Autos schneller als 100 km/h gefahren ist. Diese sind: Die Länge der Bremsspur und die Aussage eines Zeugen des Autounfalls. Es wird vereinfacht angenommen, daß jedes Beweismittel in drei von vier Fällen eine zuverlässige Beschreibung der Wirklichkeit impiziert. Die Beweismittel werden als unabhängig voneinander angenommen. Als ihren gemeinsamen Wahrscheinlichkeitswert gibt Ekelöf eine Wahrscheinlichkeit von $15/16$ an¹⁾

Ekelöfs Formel lautet offensichtlich²⁾ wie folgt:

$$w = p + q - pq$$

(p und q sind Wahrscheinlichkeiten für zwei (voneinander physikalisch unabhängige) Beweismittel).

Im obigen Beispiel gilt somit: $3/4 + 3/4 - 9/16 = 15/16$.

Wie wir sehen, ist Ekelöfs Prinzip eine Art Additionsprinzip (vgl. S.17 dieser Arbeit). Das Additionsprinzip besagt, daß die Wahrscheinlichkeit für das Vorliegen von mindestens einem Beweismittel

1) Vgl. S.58: "At least if the value of each evidentiary fact is independent of that of the other, their combined value must be greater than 3/4. But how much greater? The length of the braking marks proves that the speed exceeds 60 miles per hour in 12 out of 16 similar cases; at the same time, this is proved by the witness's statement in 3 of the 4 remaining cases. The convicting force of the combined evidentiary facts would thus be $15/16$ ". (Hervorhebung von mir). Vgl. auch: "If the convincing force of the tracks and of the statement of the witness had been only 1 out of 10, the same method of calculation would result in a combined force amounting to no more than 19 out of 100".

2) Ekelöf versäumt leider eine Formel anzugeben. Erst bei Cohen (S.99), der sich dabei auf Edman bezieht, habe ich das der Ekelöfschen Berechnung zugrundeliegende Prinzip (s.o.) gefunden; M.Eddman, Adding Independent Pieces of Evidence, in: J.Hintikka et al., Modality, Morality and other Problems of Sense and Nonsense: Essays dedicated to Sören Haldén (1973), S.180-8. Das oben in der Fn.1. angegebene Zitat läßt allerdings vermuten, daß der Ekelöfschen Berechnung auch ein anderes, viel elementarerer Prinzip zugrunde liegen könnte; vgl. "in 12 out of 16 similar cases ... in 3 of the remaining 4 cases". D.h. Ekelöf addiert, um das Ergebnis $15/16$ zu erhalten, einfach zu 12 (aus der Gesamtheit von 16) 3 (aus dem Rest 4), was 15 ergibt und relativiert diese Anzahl auf die Gesamtheit aller Fälle (also 16); das ergibt insgesamt $15/16$. Ebenso im zweiten Fall: 10 (aus der Gesamtheit von 100) + 9 (aus dem Rest $100-10=90$); das ergibt 19; die Relativierung auf die Gesamtzahl von 100 ergibt schließlich $19/100$ (das gleiche Ergebnis erhält man mit der obigen Formel).

tel gleich der Summe der Wahrscheinlichkeiten für die einzelnen Beweismittel minus die Wahrscheinlichkeit für das gemeinsame Vorliegen dieser Beweismittel ist. Die Anwendung der obigen Formel bedeutet für das Ekelöfsche Beispiel, daß in 15 von 16 Fällen mindestens eines der beiden Beweismittel vorliegt.

Ekelöfs Absicht ist vermutlich jedoch nicht, die Wahrscheinlichkeit für das Vorliegen von mindestens einem einzigen Beweismittel zu erhalten, sondern eine solche - wie er schreibt¹⁾ - für das gemeinsame Vorliegen aller Beweismittel. Dazu ist aber das Additionsprinzip nicht mehr geeignet. Stattdessen muß man das Multiplikationsprinzip (vgl. S. 16 dieser Arbeit) anwenden. Nach dem Multiplikationsprinzip ist die gemeinsame Wahrscheinlichkeit für die beiden Beweismittel (die Länge der Bremsspur und die Zeugenaussage) im obigen Beispiel aber nur $9/16$ (ihre Unabhängigkeit vorausgesetzt).

Man kann sich allerdings weiter fragen, ob Ekelöf tatsächlich, wie er behauptet, die Gesamtwahrscheinlichkeit für das Vorliegen aller Beweismittel ("combined value of the evidentiary facts") will, oder vielmehr die Wahrscheinlichkeit des Sachverhalts aufgrund des gemeinsamen Vorliegens mehrerer Beweismittel²⁾. Wenn das letztere der Fall ist, so ist weder die Anwendung des Additionsprinzips noch eine solche des Multiplikationsprinzips zulässig.

Seien H im obigen Beispiel die Hypothese, daß das Auto schneller als 100 km/h gefahren ist, A das erste Beweismittel (die Länge der Bremsspur) und B das zweite Beweismittel (die Zeugenaussage). Dann verlangt, falls wir die zweite Interpretation der Ekelöfschen Absicht annehmen, sein Prinzip, daß man die Wahrscheinlichkeit (nach der Formel $P(H/A) + P(H/B) - P(H/A)P(H/B)$)³⁾ dafür berechnet, daß

1) Vgl. S. 89, Fn. 1.

2) Darauf lassen v. a. die Ekelöfschen Formulierungen an manchen Stellen schließen; vgl. "The length of the braking marks proves that the speed exceeded 60 miles per hour...", "...this (the speed - Anm. von mir) is proved by the witness's statement" (siehe das Zitat auf S. 89, Fn. 1.). Außerdem widerspricht die Annahme des Multiplikationsprinzips der Intention Ekelöfs, der eine höhere Wahrscheinlichkeit für das gemeinsame Vorliegen aller Beweismittel als eine solche für jedes einzelne Beweismittel erreichen will; vgl. S. 57f: "...the degree of probability is increased by several concurring evidentiary facts, each of which is in itself insufficient to prove the factum probandum concerned".

3) Diese Formel ergibt sich durch Einsetzen von $P(H/A)$ für p und $P(H/B)$ für q in die Formel auf S. 89 dieser Arbeit.

Die obige Formel sei an einem einfachen Beispiel veranschaulicht: Die Wahrscheinlichkeit, mit einem Würfel eine sechs zu würfeln, also $P(H/A)$, ist $1/6$; dasselbe gilt für einen zweiten Würfel, d.h. $P(H/B)$ ist auch $1/6$. (A steht für den ersten, B für den zweiten Würfel und H ist die Hypothese, daß eine sechs gewürfelt wird). $11/36 = P(H/A) + P(H/B) - P(H/A)P(H/B)$ ist dann die Wahrscheinlichkeit dafür, daß mindestens mit einem der beiden Würfel eine sechs gewürfelt wird.

mindestens eines der beiden Beweismittel die Hypothese bestätigt¹⁾, nicht dagegen die eigentlich erwünschte Wahrscheinlichkeit der Hypothese aufgrund des gemeinsamen Vorliegens der beiden Beweismittel (das letztere läßt sich nicht nach der obigen Formel bestimmen).

Will man im obigen Beispiel also die Wahrscheinlichkeit für die Hypothese bestimmen, daß das Auto mit einer Geschwindigkeit größer als 100 km/h gefahren ist aufgrund der Länge der Bremsspur und der Zeugenaussage (d.h. $P(H/A \wedge B)$), so muß man entweder die Formel für die bedingte Wahrscheinlichkeit oder das Bayes-Theorem benutzen, keineswegs aber das Ekelöfsche (Additions-)Prinzip.

Aus dem obigen Beispiel ersieht man, daß Ekelöf den frequentistischen Standpunkt vertritt. Die frequentistische Auffassung wird jedoch von ihm nicht konsequent beibehalten. So spricht er vielerorts davon, daß wir die relativen Häufigkeiten meistens nicht feststellen können und es daher mit Schätzungen zu tun haben²⁾

Abgesehen von der Kritik am frequentistischen Ansatz kann man Ekelöf insgesamt vorwerfen, daß er von einem Prinzip (offensichtlich dem allgemeinen Additionsprinzip, welches die Wahrscheinlichkeit für das Vorliegen von wenigstens einem Beweismittel (das die Hypothese stützt) in einer Anzahl von Fällen angibt) ausgeht, das falsch ist, sowohl, wenn es darum geht, die gemeinsame Wahrscheinlichkeit mehrerer Beweismittel (welche die Hypothese stützen) zu berechnen (diese ist – im Gegensatz zur Behauptung Ekelöfs – immer kleiner als die Wahrscheinlichkeit für ein einzelnes Beweismittel), als auch, wenn es darauf ankommt, statt der Gesamtwahrscheinlichkeit für die Beweismittel eine solche für den Sachverhalt aufgrund dieser Beweismittel zu bestimmen, und daß er außerdem nicht die Mühe auf sich nimmt, dieses Prinzip zu erklären, geschweige denn zu begründen.

Und nicht zuletzt läßt sich auch die Auffassung Ekelöfs kritisieren, wonach die Urteilsbildung im Prozeß nicht auf der richterlichen Überzeugung, sondern einzig und allein auf den Tatsachen beruhen soll (vgl.S.66). Denn eine korrekte Urteilsbildung beruht weder auf der richterlichen Überzeugung allein noch auf den Tatsachen allein, sondern auf der richterlichen Überzeugung, der die Tatsachen zugrunde liegen.

1) So kann man die Wahrscheinlichkeit 15/16 im obigen Beispiel wie folgt interpretieren: In 15 von 16 Fällen wird die Hypothese von wenigstens einem der beiden Beweismittel bestätigt.

2) Vgl.S.54: "In fact, any evaluation of evidence is likely to be based, to a greater or smaller extent, upon the kind of vague statements about frequency relations..."

1.3. Kingston

In einer Reihe von Aufsätzen¹⁾ aus den Jahren 1964 bis 1966 geht Kingston auf die Anwendung statistischer Methoden in der kriminologischen Forschung, insbesondere bei der Beurteilung empirischer Befunde, ein²⁾

In seinem Aufsatz, *Application of Probability Theory in Criminalistics*, schlägt er vier Modelle vor, wie man statistisch die Identität der Herkunft (*identity of origin*) von am Tatort hinterlassenen Spuren, wie Fingerabdrücken, Blut, Haaren Fasern usw. bestimmen kann. Da dieses Thema zwar auch in den Bereich der Anwendung der Wahrscheinlichkeitstheorie (genauer: der Statistik) auf die Jurisprudenz gehört, andererseits mit dem Thema dieser Arbeit nicht unmittelbar zu tun hat, sei hier darauf nur kurz eingegangen. (Die ausführliche Schilderung der vier Modelle findet der Leser im oben genannten Aufsatz)³⁾

Wir bedienen uns in Anlehnung an Kingston der mengentheoretischen Sprache. Identifizierung (von Objekten oder Personen) wird als die Bestimmung einer Menge definiert, zu der das Objekt, das bestimmte Merkmale besitzt, gehört, bzw. als die Bestimmung der Zugehörigkeit des fraglichen Objekts zu dieser Menge.

Kingston befaßt sich mit den sog. partiell übertragbaren empirischen Befunden (Kingstons Abkürzung dafür PTE). Dies sind z.B. Fingerabdrücke, Blutspuren, Haare, Fasern, Glassplitter, welche sowohl am Tatort wie auch am Täter oder an einem Gegenstand, der mit dem Täter möglicherweise in Verbindung steht, festgestellt werden können. Wenn das der Fall ist, so kann man aufgrund des Vergleichs der Merkmale, die einerseits den Verdächtigen und andererseits die Szene betreffen, bestimmte Schlußfolgerungen bzgl. der Identität der Herkunft dieser Befunde ziehen.

Das Identifizierungsproblem lautet somit wie folgt: Gegeben ist eine am Tatort hinterlassene Spur; wie wahrscheinlich ist es, daß ein bestimmter Verdächtiger diese Spur hinterlassen hat?

1) C.R.Kingston & P.-L.Kirk, *The Use of Statistics in Criminalistics*, in: 55 *Journal of Criminal Law, Criminology and Police Science* (1964) S.514-21; C.R.Kingston, *Applications of Probability Theory in Criminalistics*, in: 60 *Journal of American Stat. Ass.* (1965), S.70-80, S.1028-1034; C.R.Kingston, *Probability and Legal Proceedings*, in: 57 *Journal of Criminal Law, Criminology and Police Science* (1966), S.93-98.

2) Auf die Wichtigkeit einer solchen Anwendung haben bereits S.D. Poisson (1837) und J.Bentham (1825) hingewiesen. Siehe auch E.B.Mode, *Probability and Criminalistics*, in: 58 *Journal of Am.Stat.Ass.* (1963) S.628-40.

3) Für den Leser, der mit der Statistik nicht vertraut ist, empfiehlt es sich vielleicht, zuerst die Ausführung zu Cullison (S.116ff dieser Arbeit) bzw. Cullison selbst (vgl. S. 79 Fn.1.) anzuschauen. Dieser bietet in seinem Aufsatz eine ausführlichere Darstellung der Binomialverteilung und ihrer Approximation (als hier bzw. bei Kingston). Diese Arbeit ist auch deswegen lesenswert, weil Cullison eine leicht verständliche Einführung in die elementare Wahrscheinlichkeitsrechnung vermittelt.

Die Menge der Objekte oder Personen, die ein bestimmtes Merkmal besitzen (Abk. ID), kann durch eine Zufallsvariable charakterisiert werden. Es ist möglich, die Wahrscheinlichkeitsfunktion für diese Zufallsvariable zu bestimmen. Wie das funktioniert, wird von Kingston in den ersten beiden Modellen veranschaulicht. Sie unterscheiden sich voneinander nur dadurch, daß im ersten angenommen wird, daß der Angeklagte aus der Menge der Verdächtigen zufällig herausgegriffen wird, während im zweiten diese (nicht ganz realistische) Annahme nicht mehr gilt. Die detaillierte Analyse der Modelle ist für den Fall durchgeführt, daß die geschätzte Wahrscheinlichkeitsverteilung binomial mit einem Erwartungswert nicht größer als 1 ist. Es wird außerdem gezeigt, daß die Annahme der Zufälligkeit der Wahl des Verdächtigen um so unwesentlicher wird, je kleiner der Erwartungswert ist.

Gegeben sei eine Wahrscheinlichkeitsfunktion $p(x)$, die für die Zufallsvariable X bestimmt wird, wobei X den Wert x annimmt (X ist die Anzahl der Elemente in der Menge ID).

1. Modell:

Es gilt die Annahme, daß die Wahl des Verdächtigen eine Zufallswahl aus der Menge ID ist.

n sei die Anzahl der Elemente einer Bezugsmenge N , hinsichtlich derer man die Stichprobe vornimmt. Es wird angenommen, daß diese Elemente ihrerseits eine zufällige Stichprobe aus irgendeiner unerschöpflichen Gesamtheit G sind. Die Bezugsmenge N (die n Elemente enthält) ist somit nur eine Zusammensetzung aus vielen möglichen Stichproben aus G .

Die Wahrscheinlichkeit, einen Verdächtigen (mit den in Frage kommenden Merkmalen) aus G zu wählen, sei ϕ . Weiter sei angenommen, daß die Anzahl der Verdächtigen in N unbekannt ist. Die Anzahl der Verdächtigen, die eine beliebige Menge aus n Elementen enthält, ist eine Zufallsvariable X , die nach $b(x; n, \phi)$ binomial verteilt ist.

p' sei die Schätzung von ϕ und $p(x)$ die binomiale Verteilung:

$$p(x) = \binom{n}{x} (p')^x (1-p')^{n-x}$$

Da nach Voraussetzung mindestens ein Verdächtiger existieren muß, gilt $x \geq 1$. Die Wahrscheinlichkeit, daß der Verdächtige korrekt gewählt (identifiziert) wurde, lautet somit:

$$E(1/X | 1 \leq x \leq n) = \frac{\sum_{x=1}^n (1/x) p(x)}{\sum_{x=1}^n p(x)}$$

$1 - E(1/X | 1 \leq x \leq n)$, oder $1 - E$, ist die Wahrscheinlichkeit, daß der Verdächtige falsch identifiziert wurde (Irrtum in der Wahl des 'origins').

Zur Veranschaulichung: Es sei ein Fingerabdruck am Tatort sichergestellt. Ferner sei bekannt (wie man dazu kam, sei hier irrelevant), daß die Wahrscheinlichkeit dafür, daß irgendeine Person die Merkmale dieses Fingerabdrucks aufweist, ca. $1/(3 \times 10^9)$ ist. Durch eine Routineüberprüfung entdeckt man eine Person, die den gleichen Fingerabdruck hat. Wie groß ist dann die Wahrscheinlichkeit für eine irrtümliche Identifizierung (bzw. Beschuldigung), wenn der Fingerabdruck der einzige Befund ist?

Min $n = 3 \times 10^9$ ($\approx$ die Hälfte der Weltbevölkerung) und $p' = 1/(3 \times 10^9)$ ist der Erwartungswert von X $np' = 1$. In diesem Fall können wir die Funktion $p(x)$ durch eine Poissonverteilung mit dem Parameter $\lambda = 1$ approximieren. Es gilt somit:

$$E(1/X | 1 \leq x \leq n) = \frac{\sum_{x=1}^{\infty} 1/(x \cdot x!)}{\sum_{x=1}^{\infty} 1/x!} = \frac{1}{e-1} \cdot \sum_{x=1}^{\infty} \frac{1}{x \cdot x!}$$

$$\approx 0.77.$$

(e ist die irrationale Zahl 2.718..., die Basis des natürlichen Logarithmus).

Es folgt, daß die Wahrscheinlichkeit des Irrtums, 1-E, ungefähr 0.23 ist.

mit $np' = \lambda$ und $\lambda \leq 1$ lautet die Formel für genügend große n:

$$E(1/X | 1 \leq x \leq n) = \frac{1}{e^{\lambda} - 1} \cdot \sum_{x=1}^{\infty} \frac{\lambda^x}{x \cdot x!}$$

(Im allgemeinen kann man zeigen, daß 1-E um so kleiner wird, je kleiner λ ist; z.B. für $\lambda = 1/10^2$ ist 1-E = 0.00249).

2.Modell:

Hier wird die Annahme fallengelassen, daß der Verdächtige zufällig aus der Menge ID gewählt wird. In diesem Fall müssen wir die Wahrscheinlichkeit bestimmen, daß ein weiteres Element der Menge ID gefunden werden kann, oder äquivalent dazu, daß ein Irrtum in der Identifizierung des Verdächtigen auf der Basis der PTE möglich ist. $P(\text{Er}|X=x)$ sei diese Wahrscheinlichkeit (sie hängt vom Wert von X ab). Es gilt:

$$P(\text{Er}|X=x) = 0 \quad \text{für } x = 1,$$

$$P(\text{Er}|X=x) = 1 \quad \text{für } x > 1.$$

Der Erwartungswert von $P(\text{Er}|X=x)$ ist die Wahrscheinlichkeit eines Irrtums (wobei die Bedingung bzgl. x fallengelassen wird). Sei $\lambda = np' \leq 1$. Dann ist X approximativ poissonverteilt mit λ als Parameter. Für $x \geq 1$ lautet die Formel:

$$P(\text{Er}) = \frac{\sum_{x=2}^{\infty} p(x)}{\sum_{x=1}^{\infty} p(x)} = 1 - \frac{\lambda}{e^{\lambda} - 1}$$

Für $\lambda = 1$ ist $P(\text{Er}) = 0.41803$, für $\lambda = 1/10^2$ ist $P(\text{Er}) = 0.00500$.

Die Werte von $P(\text{Er})$ sind ungefähr doppelt so hoch wie die Werte für 1-E im ersten Modell. Je kleiner λ ist, desto geringer ist die Bedeutung der Differenz zwischen $P(\text{Er})$ und 1-E bei der Beurteilung des Befundes (so daß das 1.Modell gegenüber Verletzungen der Annahme der Zufälligkeit bei der Wahl des Verdächtigen aus ID widerstandsfähiger wird).

3.Modell:

Dies ist ein allgemeines Modell, das weniger restriktiv ist als die zwei vorhergehenden.

Um es zu veranschaulichen, nehmen wir folgende hypothetische Situation an: C sei eine Bevölkerungsgruppe, die n+1 Mitglieder aufweist (C kann ihrerseits zufällig aus einer noch größeren (unerschöpflichen) Population S1 gewählt werden).

Nehmen wir an, daß der Täter blondes Haar hat; die Wahrscheinlichkeit für das Vorkommen blonden Haars in S1 sei p_1 .

Weiter sei angenommen, daß verschiedene Befunde (aus einer Gesamtheit S2 aller möglichen Befunde) die Tatortsituation kennzeichnen, wie etwa ein zerbrochenes Fenster, Metallspäne vom Gebrauch der Einbruchsinstrumente usw. (Dabei ist wich-

tig, daß es sich hier um Befunde, die man leicht vom Tatort wegtragen kann, d.h. um PTE, handelt). Es ist möglich, daß der Täter (wir nennen ihn kurz B) einen oder mehrere solcher Befunde vom Tatort an sich (in seiner Kleidung, Wohnung, Auto etc.) aufweist. Wir wissen jedoch nicht, welche. Sei p_3 die Wahrscheinlichkeit dafür, daß B ein bestimmtes Merkmal aus der (unerschöpflichen) Gesamtheit S_3 aller am Tatort möglichen PTE-Merkmale und ihrer Kombinationen an sich hat (z.B. Glassplitter vom zerbrochenen Fenster). p_2 sei die Wahrscheinlichkeit für das Auffinden gleicher Glassplitter wie am Tatort an einer Person in der Gesamtheit S_2 .

Es sei nun ein Verdächtiger mit blondem Haar gefunden und es seien bei der nachfolgenden Untersuchung an seiner Kleidung Glassplitter entdeckt worden. Uns interessiert die Frage, wie wahrscheinlich es ist, daß der Verdächtige aufgrund der Befunde irrtümlicherweise beschuldigt wird, der Täter zu sein. (Nur blondes Haar und Glassplitter an der Kleidung sollen als Beweismittel verwendet werden).

Es sind vier erschöpfende und sich wechselseitig ausschließende Ereignisse möglich ($\mathcal{P}$ sei die Abkürzung für $C \setminus B$)¹⁾:

E_1 : B weist Glassplitter auf; $\mathcal{P}$ enthält keine blonden Personen mit Glassplittern;
 E_2 : B weist Glassplitter auf, $\mathcal{P}$ enthält eine oder mehrere blonde Personen mit Glassplittern;

E_3 : B weist keine Glassplitter auf; $\mathcal{P}$ enthält keine blonden Personen mit Glassplittern;

E_4 : B weist keine Glassplitter auf; $\mathcal{P}$ enthält eine oder mehrere Personen mit Glassplittern.

Nach Voraussetzung ist E_3 nicht möglich (da der Verdächtige Glassplitter an seiner Kleidung aufweist); deswegen gilt:

$$k(P(E_1) + P(E_2) + P(E_4)) = 1.$$

(k ist dabei eine Normierungskonstante und $P(E_i)$ die Wahrscheinlichkeit des Eintretens des Ereignisses E_i , $i = 1, 2, 4$).

Es folgt:

$$k = \frac{1}{1 - P(E_3)}.$$

Ein Irrtum in der Folgerung, daß es sich bei dem gefundenen Verdächtigen um den Täter handelt, ist möglich, wenn entweder E_2 oder E_4 eintritt. Es gilt somit:

$$P(E_T) = k \cdot P(E_2) + k \cdot P(E_4) = 1 - k \cdot P(E_1).$$

Die Zahl der blonden Personen in $\mathcal{P}$ ist eine Zufallsvariable Y und nach einer Binomialverteilung mit den Parametern n und p_1 verteilt; auch die Zahl der Personen in $\mathcal{P}$, die Glassplitter aufweisen, ist eine Zufallsvariable Z , die eine Binomialverteilung mit den Parametern n und p_2 besitzt. Und schließlich ist die Zahl der blonden Personen in $\mathcal{P}$, die Glassplitter aufweisen, wiederum eine Zufallsvariable X , die einer Binomialverteilung mit den Parametern n und $p_1 p_2 = p^*$ gehorcht (es wird angenommen, daß das Auffinden der Glassplitter aus S_2 und eines blonden Subjekts in S_1 voneinander unabhängig sind).

Unter der Annahme, daß die Poissonapproximation der Binomialverteilung möglich ist, folgt für die Wahrscheinlichkeit dafür, daß $\mathcal{P}$ keine blonde Person mit Glassplittern enthält, $\approx e^{-\lambda}$, mit $\lambda = np^*$. Die Wahrscheinlichkeit, daß B keine Glassplitter aufweist, ist $(1-p_2)(1-p_3) = q'$. Die Wahrscheinlichkeit, daß B Glassplitter aufweist, ergibt sich somit als $1 - q' = p'$.

Mit der Unabhängigkeitsannahme der beiden Ereignisse folgt daraus:

1) zu lesen: "C ohne B".

$$k = \frac{e^\lambda}{e^\lambda - 1 + p'}$$

und

$$P(E_1) = \frac{p'}{e^\lambda}$$

Insgesamt erhalten wir schließlich:

$$P(Er) = \frac{e^\lambda - 1}{e^\lambda - 1 + p'} = \frac{\lambda}{\lambda + p' \cdot \left(\frac{\lambda}{e^\lambda - 1} \right)} \quad (*)$$

4. Bayes-Modell:

Dieses Modell wird von Kingston als Approximation für die im dritten Modell errechnete Irrtumswahrscheinlichkeit vorgeschlagen (unter der Voraussetzung, daß diese Wahrscheinlichkeit klein genug ist).

Die Apriori-Wahrscheinlichkeit für das Auffinden des Täters wird als $1/(n+1)$ angesetzt ($n+1$ ist die Anzahl der Mitglieder in C; vgl. 3. Modell); d.h. man geht, bevor die Befunde beurteilt werden, davon aus, daß der Verdächtige zufällig aus C gewählt werden kann.

Nun wird die Wahrscheinlichkeit dafür gesucht, daß der Verdächtige aufgrund der empirischen Befunde nicht der Täter ist; Abk. dafür ist $P(\bar{B}/Ev)$, die Abk. für das Komplement ist entsprechend $P(B/Ev)$.

Die Bayes-Formel lautet in diesem Fall wie folgt:

$$P(\bar{B}/Ev) = \frac{P(Ev/\bar{B})P(\bar{B})}{P(Ev/\bar{B})P(\bar{B}) + P(Ev/B)P(B)}$$

Mit der Notation aus dem 3. Modell, $P(Ev/\bar{B}) = p^*$ und $P(Ev/B) = p'$, ist

$$P(\bar{B}/Ev) = \frac{\frac{np^*}{n+1}}{\frac{np^*}{n+1} + \frac{p'}{n+1}} = \frac{\lambda}{\lambda + p'} \quad (**)$$

Falls λ klein ist, ist $(\lambda/(e^\lambda - 1))$ nahe an 1, und die beiden Formeln (*) und (**) liefern approximativ die gleichen Resultate.

Die Formel (**) kann auch so geschrieben werden:

$$P(\bar{B}/Ev) = \frac{n}{n + \frac{p'}{p^*}}$$

Neben n müssen wir somit nur das Verhältnis der zwei Wahrscheinlichkeiten p' und p^* kennen.

Kritik an den Modellen Kingstons:

1. Die erste Kritik betrifft die exakte Bestimmung von p' (in den beiden ersten Modellen) bzw. p_1, p_2 und p_3 (in den beiden letzten Modellen). p' (ebenso $p_i, i = 1, 2, 3$) wird in bezug auf eine Gesamtheit angegeben, deren Umfang und deren Verteilung bzgl. des fraglichen Merkmals völlig unbekannt bleiben. Wie p' (und p_i) zustande kommt, wird nicht erläutert. (Es ist nicht übertrieben zu behaupten, daß p' (ebenso p_i) in den meisten Fällen nicht einmal schätzungsweise angebbbar, geschweige denn exakt bestimmbar ist).
 2. Die Bestimmung der Wahrscheinlichkeit für die Täterschaft erfolgt lediglich aufgrund des statistischen Datums. Es ist klar, daß das statistische Datum nur einen Bruchteil dessen ausmacht, was im Prozeß vorgelegt wird. Außerdem soll man nicht vergessen, daß das statistische Datum nicht stärker sein kann als die Aussage (die Glaubwürdigkeit) eines Zeugen bzw. eines Experten, auf der das statistische Datum aufbaut.
- Bei dem statistischen Ansatz fehlt somit die Berücksichtigung des nichtstatistischen Datums bei der wahrscheinlichkeitstheoretischen Beurteilung des in Frage kommenden Sachverhalts¹⁾.

Ansonsten zeigen die Modelle von Kingston die (allerdings weniger praktische, als theoretische) Möglichkeit, bei einer Art von empirischen Befunden die Statistik in der kriminologischen Forschung sinnvoll zu verwenden. (Voraussetzung ist jedoch, daß es sich bei diesen Befunden um PTE (partiell übertragbare Befunde, d.h. um solche, die sowohl am Tatort wie auch beim Täter gefunden werden können) handelt. Das Vorgehen von Kingston bietet außerdem die Möglichkeit, das statistische Datum auch dann zu berücksichtigen, wenn die Anzahl der Verdächtigen (d.h. der Personen mit den bestimmten, in Frage kommenden Merkmalen) in einer Population vom Umfang n völlig unbekannt ist. Dies können wir nach Kingston durch die Binomialverteilung mit den Parametern n und ϕ bestimmen; Ist n groß und ϕ sehr klein, so läßt sich die Binomialverteilung durch die Poissonverteilung approximieren. (Daß man nicht nur mit der Wahl von ϕ , sondern auch mit der von n seine Schwierigkeiten haben wird - will man diese Formel auch praktisch anwenden - sei hier nur nebenbei erwähnt)²⁾³⁾

1) Eine Möglichkeit, neben dem statistischen auch das nichtstatistische Datum bei der wahrscheinlichkeitstheoretischen Beurteilung des Sachverhalts zu berücksichtigen, erläutern Finkelstein und Fairley in ihrem Aufsatz (vgl. S.123ff dieser Arbeit).

2) Auch J.B.Parker beschäftigt sich mit der statistischen Behandlung von kriminologischen Identifizierungsproblemen; vgl. seine Artikel: "A Statistical Treatment of Identification Problems", in: 6 For.Sci.Soc. (1966), S.33-39; "The Mathematical Evaluation of Numerical Evidence", ibid, S.134-144. Auf diese Arbeiten wird allerdings hier nicht eingegangen, denn ihr Verständnis verlangt vom Leser eine über das Anfangsstadium hinausgreifende Kenntnis der mathematischen Statistik.

3) Einige weitere schwerwiegende Einwände gegen die Anwendung der Binomialformel gibt Cullison (im Zusammenhang mit dem Fall 'People v. Collins') an; siehe S.120 dieser Arbeit.
Zur Kritik an Kingston vgl. auch R.L.Smith&R.P.Charrow, Upper und Lower Bounds for Probability of Guilt Based on Circumstantial Evidence, in 70 Journal of the American.Stat.Ass. (1975), S.555-560; darauf wird später noch (auf S.121f dieser Arbeit) kurz eingegangen.

1.4. Kegel¹⁾

Kegel unterscheidet zwischen "Voll- und Anscheinsbeweis". Der "Vollbeweis" liegt dann vor, wenn eine Tatsache "immer mit der anderen verknüpft" ist (d.h. wenn der Schluß von einer Tatsache auf eine andere Tatsache sicher ist). Ist der Schluß von der einen Tatsache auf die andere "nur wahrscheinlich"²⁾, so handelt es sich nach Kegel um einen Anscheinsbeweis. "Nur wahrscheinlich" wird dabei von Kegel im Sinn von "meist" verwendet³⁾.

Wie bekannt, erwartet die Rechtsprechung (insbesondere die deutsche), daß sowohl im Straf- wie auch im Zivilprozeß der volle Beweis (d.h. der Beweis, der die Wahrheit hinsichtlich der Tatsachen und Sicherheit hinsichtlich der Überzeugung von der Feststellung solcher Tatsachen behauptet) vorliegt⁴⁾.

Daß der Vollbeweis - entgegen der Meinung der Rechtsprechung - auch ein Beweis nach Wahrscheinlichkeiten ist, wurde bereits gezeigt⁵⁾.

Auch Kegel spricht davon, daß der volle Beweis in Wirklichkeit ein Anscheinsbeweis sei⁶⁾.

1) G. Kegel, Der Individualanscheinsbeweis und die Verteilung der Beweislast nach überwiegender Wahrscheinlichkeit, in: Festschrift für Kronstein (1967), S. 321-44.

Kegel kann als erster im deutschen Sprachraum gelten, der mit Wahrscheinlichkeitsurteilen praktisch zu arbeiten versucht.

Zur Relevanz der Wahrscheinlichkeitstheorie vgl. auch v. Kries (1889) und Weitnauer (1966); S. 63f, 79f dieser Arbeit.

2) Vgl. ibid., S. 323.

3) Vgl. ibid., S. 323: "Der Schluß von der einen Tatsache auf die andere kann aber auch nur wahrscheinlich sein...: die eine Tatsache ist meist mit der anderen verknüpft; die eine Tatsache ist Anscheinsbeweis der anderen; S. 343: "Vom Anscheinsbeweis sagt man, er beruhe auf einem "typischen Geschehensablauf". Damit meint man: unter bestimmten Umständen ist es meist, d.h. mit überwiegender Wahrscheinlichkeit so". Zum Voll- und Anscheinsbeweis vgl. auch Kegel, S. 327f.

4) Vgl. dazu 1. Kap., § 2 dieser Arbeit.

5) Vgl. ibid.

6) Vgl. Kegel, S. 328: "Dann aber herrscht in Wahrheit immer der Anscheinsbeweis: man muß sich begnügen mit dem, was unter den ermittelten Umständen meist so ist... Der Vollbeweis mit Hilfe von Natur- und Denkgesetzen (es ist notwendig, es ist immer so) kann jedenfalls nicht mehr deutlich abgegrenzt werden vom Anscheinsbeweis mit Hilfe von Erfahrungssätzen (es ist typisch, es ist meist so); und weiter S. 334: "Die andere Antwort auf das Verhältnis von Vollbeweis (volle richterliche Überzeugung = sehr hohe Wahrscheinlichkeit) und Anscheinsbeweis (überwiegende Wahrscheinlichkeit) dehnt den Anscheinsbeweis aus: wir erreichen immer nur Wahrscheinlichkeit, nie Sicherheit. Der Grad kann schwanken. Aber überwiegende Wahrscheinlichkeit genügt". (Hervorhebung im Text).

Die Erörterungen Kegels zur wahrscheinlichkeitstheoretischen Beurteilung tragen eher einen informellen Charakter. Es ist daher interessant festzustellen, ob seine Intuition den mathematischen Wahrscheinlichkeitsprinzipien gehorcht.

Dies ist in der Tat so, wie zahlreiche Äußerungen von Kegel belegen; so schreibt er auf S.325:

"Im Bündel erhöht sich die Wahrscheinlichkeit. Ein Bündel von Hilfstatsachen, die einzeln mit geringer Wahrscheinlichkeit auf eine andere Tatsache hindeuten, kann sich zum Anscheinsbeweis (meist ist es so) oder zum Vollbeweis (immer ist es so) verstärken".

Dies ist aber nichts anderes als die informelle Beschreibung dessen, was durch die Formel für die bedingte Wahrscheinlichkeit (falls das Gesamtdatum berücksichtigt wird) erreichbar ist. Die verbale Umschreibung des Negationsprinzips ist die folgende Behauptung (S.327):

"Die Wahrscheinlichkeit eines Ablaufs sinkt in dem Maße, in dem ein anderer Ablauf wahrscheinlich wird".

Auch die folgende Äußerung Kegels kann man wahrscheinlichkeitstheoretisch veranschaulichen (S.328):

"Denn nach Lage der Sache...kann das Unwahrscheinliche das Wahrscheinlichste sein, weil alles andere noch unwahrscheinlicher ist...";

etwa durch die Wahrscheinlichkeitsverteilung:

0,3, 0,1, 0,1, 0,1, 0,1, 0,1, 0,1, 0,1.

In diesem Fall haben wir 8 Möglichkeiten, die wahrscheinlichste von ihnen hat nur die Wahrscheinlichkeit von 0,3 (was relativ gering ist), die restlichen 7 haben jeweils die Wahrscheinlichkeit von 0,1, was jede der 7 Alternativen aber noch unwahrscheinlicher macht¹⁾

Interessant sind die von Kegel geschilderten juristischen Fälle, wo er zeigt, daß die Rechtsprechung, auch wenn sie das bestreitet, in Wirklichkeit in Wahrscheinlichkeitsurteilen denkt. Wir wollen hier den ersten aus den von Kegel angeführten Fällen wörtlich wiedergeben²⁾

"Ein Ehemann verklagt das Land Schleswig-Holstein auf Schadenersatz, weil seine Frau in der Medizinischen Klinik der Universität Kiel mit Lues infiziert worden sei. Die Frau hat am 25. Juni 1942 eine Bluttransfusion erhalten. Blutspender war nach Behauptung des Klägers der Steinträger L. Im Krankenblatt dieses Mannes ist zwar keine Blutspende eingetragen, wohl aber am 1. Juli 1942 eine positive

1) Die folgende Behauptung Kegels ist allerdings falsch (S.327):

"Je genauer der Sachverhalt festgestellt wird, um so wahrscheinlicher wird er", denn die Ableitung eines (hohen) Wahrscheinlichkeitsgrades für den Sachverhalt hat mit der Genauigkeit seiner Feststellung nichts zu tun.

2) Kegel, S.329 ("Lues-Fall").

Wassermannreaktion, die sich bei einer am 11. Juni 1942 angeordneten Untersuchung gezeigt hat. Als die Frau ihrerseits im Mai 1947 Blut spenden will, ist auch ihre Wassermannreaktion positiv. Dagegen treten typische Lueserscheinungen bei ihr nicht auf. Die Wassermannreaktionen ihres Mannes und ihrer beiden Söhne sind negativ.

Das OLG Schleswig weist die Klage ab. Es sieht zwar die Blutspende und eine Lues III. Stufe bei dem Steinträger als erwiesen an, dagegen weder eine Lues der Frau (die positive Wassermannreaktion könne an anderen Erkrankungen liegen) noch, wenn eine Lues vorliegen oder vorgelegen haben sollte, eine Verursachung durch die Blutspende. Auch ein Anscheinsbeweis fehle; denn Ansteckung durch Übertragung von Blut eines Lueskranken III. Stufe sei selten.

Der BGH verweist zurück. Der Anscheinsbeweis sei geführt. Man dürfe nicht nur von der Ursache (Transfusion des Blutes eines Luetikers III. Stufe) auf die Folge (Ansteckung) sehen und sie wegen ihrer Seltenheit ausscheiden. "Es ist vielmehr auch von dem festgestellten Erfolg aus rückblickend zu fragen, welche Anhaltspunkte für das Vorhandensein etwaiger anderer möglichen Ursachen bestehen. Ergibt diese Prüfung, daß bei einem bestimmten Krankheitsbild für eine Ursache feste Anhaltspunkte bestehen, die diese Ursache - wenn auch nur entfernt - als möglich erscheinen lassen, während für die anderen in Frage kommenden Ursachen solche Anhaltspunkte tatsächlicher Art völlig fehlen, so spricht der Beweis des ersten Anscheins für die erste Ursache". Im vorliegenden Falle bestehe für eine Erkrankung gerade an Lues und für ihre Verursachung gerade durch die Transfusion ein - obschon geringer - Anhalt durch die Erkrankung des Blutspenders an Lues III. Stufe. Dafür, daß die Krankheit eine andere sei und daß sie eine andere Ursache als die Transfusion habe, fehle hingegen jeder Anhalt".

Aus der Entscheidung des BGH ersieht man, daß er nach dem wahrscheinlichen Sachverhalt - Erkrankung an Lues und Verursachung durch eine Bluttransfusion - entschieden hat: Denn neben diesem Sachverhalt nimmt der BGH andere Möglichkeiten (auch wenn sie nicht empirisch bestätigt sind) an, nämlich andere (als Lues) Krankheiten oder andere Ursachen (als die einer Bluttransfusion), von denen jede jedoch weniger wahrscheinlich ist als der erste Sachverhalt. Man kann sogar vermuten, daß der BGH nicht einmal nach dem Kriterium der "überwiegenden Wahrscheinlichkeit" im Sinne "größer als 0.5" entschieden hat, sondern 'nur' nach dem wahrscheinlicheren Sachverhalt. Denn dafür sprechen gerade die suggestiven Formulierungen wie: "die diese Ursache - wenn auch nur entfernt - als möglich erscheinen lassen", "obschon geringer - Anhalt durch die Erkrankung des Blutspenders an Lues".

Während der BGH in nichtbedingten Wahrscheinlichkeitsurteilen denkt, ist das Vorgehen Kegels als ein solches in bedingten Wahrscheinlichkeiten zu charakterisieren. So schreibt er (S. 329f):

"Gesetzt, das Krankheitsbild der Frau passe außer auf Lues auf zwei weitere Krankheiten und diese kämen ebenso selten vor wie eine Lues durch Blutübertragung. Dann muß ich fragen: "Wie wahrscheinlich ist eine der beiden anderen Krankheiten, wenn außerdem Blut eines Luetikers übertragen ist?" Die Antwort lautet: "Höchst selten und darum ist nach den ermittelten Umständen eine Lues am wahrscheinlichsten".

Kegel geht hier in korrekter Weise von der Apriori-Wahrscheinlichkeit für diese drei Krankheiten (nach Voraussetzung sind sie alle gleichermaßen selten, daher können wir getrost für jede von ihnen die Apriori-Wahrscheinlichkeit gleich $1/3$ annehmen, vorausgesetzt

daß diese drei Krankheiten den Möglichkeitsraum erschöpfen) über zu der Aposteriori Wahrscheinlichkeit, die durch die Information bedingt ist, daß das Blut eines Luetikers übertragen wurde. Anhand dieser zusätzlichen Information leitet Kegel eine höhere Wahrscheinlichkeit für die Lues als für jede der beiden anderen Krankheiten ab.

Würdigung

Kegels Beschäftigung mit der Wahrscheinlichkeit ist nicht nur deshalb wichtig, weil mit Kegel erstmals in Deutschland der Vollbeweis als Wahrscheinlichkeitsbeweis erkannt wurde, sondern weil sie auch den ersten Versuch darstellt, die Regel "nach überwiegender Wahrscheinlichkeit"¹⁾ in die deutsche Rechtsprechung einzuführen und zu begründen. So ist sie nach Kegel "das Beste, was wir haben können", sie genüge immer, sie garantiere allein die Gerechtigkeit usw.²⁾ Im Übrigen ist Kegels Intuition richtig und nur durch die folgende Formulierung zu ergänzen: sie ist auch die einzige, die rational (bzw. formal) gerechtfertigt werden kann³⁾

Was die Interpretation und Bestimmung der Wahrscheinlichkeit angeht, so bezieht sich Kegel auf den Vorschlag Weitnauers⁴⁾, dabei die Statistik zu verwenden. Daß er jedoch nicht ganz konsequent den frequentistischen Standpunkt vertritt, belegt die Tatsache, daß er

1) Vgl. Kegel, S. 334: "Überwiegende Wahrscheinlichkeit heißt: mindestens etwas mehr als 50% Wahrscheinlichkeit".

2) Vgl. *ibid.*, S. 334, 335, 343, 344. S. 343f: "Anscheins- und Individualanscheinsbeweis beruhen auf überwiegender Wahrscheinlichkeit, die sich aus den typischen oder besonderen Umständen des Falls ergibt. Vom Vollbeweis sagt die h.M., er begründe "volle richterliche Überzeugung", d.h. zwar nicht Sicherheit, aber doch mehr als bloß überwiegende Wahrscheinlichkeit, nämlich sehr hohe Wahrscheinlichkeit. Wie kommt es, daß einmal überwiegende Wahrscheinlichkeit genügt (Anscheins- und Individualanscheinsbeweis), das andere Mal sehr hohe Wahrscheinlichkeit nötig sein soll (Vollbeweis)? Die Antwort: es gibt (von besonderen Gründen der Beweislastverteilung abgesehen) nur einen Beweis, die überwiegende Wahrscheinlichkeit genügt immer".

"Der Grund dafür, daß überwiegende Wahrscheinlichkeit zum Beweis ausreicht, liegt in der Gerechtigkeit: wenn wir schon nicht sicher oder fast sicher sein können, geben wir immer noch besser dem recht, der wahrscheinlich recht hat, als dem, der wahrscheinlich unrecht hat". (Hervorhebung im Text).

3) Dazu siehe den II. Teil dieser Arbeit.

4) Vgl. Kegel, S. 338 (Weitnauer, vgl. S. 63f dieser Arbeit).

von Schätzungen und Ungenauigkeiten spricht¹⁾ die ihn schließlich zur Äußerung verleiten "Die genaue Bestimmung des Wahrscheinlichkeitsgrades ist zum Glück meist überflüssig"²⁾ Eine solche Äußerung wäre nicht möglich gewesen, wenn Kegel den subjektivistischen Ansatz gekannt hätte, denn danach ist die numerische Bestimmung des Wahrscheinlichkeitsgrades nicht ungenauer bzw. verschwommener als irgendeine sonstige subjektive Beurteilung des Sachverhalts.

1) Kegel, S.338f. Interessant ist in diesem Zusammenhang, daß Kegel gerade der Ungenauigkeit in der Bestimmung der Wahrscheinlichkeitsgrade die Verantwortung dafür gibt, daß die Formel von der "vollen richterlichen Überzeugung" sich so lange halten konnte; vgl.S.339: "Man kann z.B. die Grenzlinie der "vollen richterlichen Überzeugung" oder der überwiegenden Wahrscheinlichkeit als nicht erreicht ansehen, wenn der nötige Wahrscheinlichkeitsgrad nicht genau zu bestimmen ist. Doch dürfte das nicht praktisch sein, weil die tatsächliche Feststellung oft, wenn nicht sogar meist, keine genauen Wahrscheinlichkeitsgrade liefert. Nur diese Verschwommenheit dürfte überhaupt erklären, daß sich die Formel der "vollen richterlichen Überzeugung" bis heute hat halten können, ohne durch ihre übermäßige Strenge Schaden anzurichten".

2) *ibid*, S.344 (Hervorhebung im Text).

1.5.Schreiber¹⁾

Schreiber tritt mit dem Anspruch auf, eine Theorie zu formulieren, die Regeln für die richterliche Beweiswürdigung angibt²⁾

Abgesehen davon, daß Schreiber zu Recht das Evidenzgefühl des Richters als Wahrheitskriterium ablehnt, da es nicht intersubjektiv nachprüfbar ist,³⁾ ist er nichtsdestotrotz der Meinung, daß die Wahrheit in Rechtsfällen feststellbar sei, wobei dies aufgrund des "empirischen Wahrheitskriteriums" geschehe⁴⁾

Schreiber legt seinen Ausführungen einen objektivistischen Ansatz zugrunde. So definiert er den Beweiswert von Zeugenaussagen⁵⁾ als "Wahrscheinlichkeit der Wahrheit, daß die Aussage des Beweismittels mit der Wirklichkeit übereinstimmt"⁶⁾ diese "wiederum als das Verhältnis der Anzahl der wahren Aussagen zu der Anzahl der wahren und falschen Aussagen"⁷⁾

Mit Schreibers Worten (S.20):

1) R.Schreiber, Theorie des Beweiswertes für Beweismittel im Zivilprozeß, Berlin-Heidelberg-New York (1968).

2) *ibid.*, S.3.

3) *ibid.*, S.6, 40. Richtig ist auch die Auffassung Schreibers, daß es sich bei der Tatsachenfeststellung nicht um eine Rechtsfrage, sondern um eine empirische Frage handelt, und sie daher nicht mit rechtlichen, sondern mit empirischen Methoden geklärt werden muß (vgl. Schreiber, S.7).

4) Vgl. *ibid.*, S.9: "Der eindeutige Vorzug des empirischen Wahrheitskriteriums ist, daß es erstens die Entscheidung über die Wahrheit oder Falschheit von Behauptungen der subjektiven Meinung des Richters entzieht..." Schreiber übersieht anscheinend, daß das empirische Wahrheitskriterium (=Verifikationskriterium) bei der von ihm angeschnittenen Fragestellung - der Wahrscheinlichkeitstheoretischen Beurteilung von Beweismitteln - nicht angemessen ist.

5) Für Schreiber stellt der Zeugenbeweis das schwierigste juristische Problem überhaupt dar (vgl.S.17). Mißverständlicherweise wählt er für den - von ihm objektivistisch gemeinten - Ausdruck "Beweiswert der Zeugenaussage" die subjektivistisch gefärbte Bezeichnung "Glaubwürdigkeit".

6) Schreiber, S.25. Nach Schreiber lassen sich Zeugenaussagen intersubjektiv nachprüfen, indem man sie durch den Augenschein verifizieren kann (vgl. S.19). Wenn das so ist, kann man mit Nell fragen: Wozu dann die Zeugenaussage? Der Augenschein genügt! (vgl. Nell, S.105, der diesen Gedanken etwas anders formuliert).

Es ist eine triviale Feststellung, wenn man sagt, daß gerade die Nichtnachprüfbarkeit der Zeugenaussage die eigentliche Schwierigkeit des Zeugenbeweises darstellt.

7) *ibid.*, S.25f.

"Der Begriff der Wahrscheinlichkeit ist auch auf Zeugenaussagen anwendbar, weil er auf Mengen Bezug nimmt. Die Wahrscheinlichkeit der Wahrheit einer Zeugenaussage ist dann erklärt als das Verhältnis der Anzahl der wahren Zeugenaussagen, die durch diese Umstände gekennzeichnet sind, zu der Anzahl der wahren und falschen Zeugenaussagen, die durch die angegebenen Umstände bestimmt sind. Die Aussagefähigkeit dieses Maßstabes der Glaubwürdigkeit hängt demnach davon ab, ob sich die Wahrheit der Aussage gewissen Umständen dieser Aussage zuordnen läßt. Da diese eine Voraussetzung für jede sinnvolle Beurteilung der Glaubwürdigkeit einer Zeugenaussage ist, entscheidet diese Frage nicht nur über den Wahrscheinlichkeitsbegriff der Glaubwürdigkeit, sondern über jede Bewertung von Zeugenaussagen überhaupt. Da die Möglichkeit einer Bewertung von Zeugenaussagen nicht in Frage gestellt wird, ist demnach auch die Voraussetzung für den Wahrscheinlichkeitsbegriff gegeben".

Neben der Ungenauigkeit der Wendung "Wahrscheinlichkeitsbegriff der Glaubwürdigkeit", was nichts anderes bedeutet als ein "Wahrscheinlichkeitsbegriff der Wahrscheinlichkeit"¹⁾ - d.h. ein von Schreiber nicht intendierter iterierter Wahrscheinlichkeitsbegriff -, ist die oben zitierte Stelle auch in anderen Hinsichten schwer zu verstehen. Es ist nämlich nicht klar, wie der Satz gemeint ist

"Die Wahrscheinlichkeit der Wahrheit eines Zeugenaussage ist dann erklärt als das Verhältnis der Anzahl der wahren Zeugenaussagen, die durch diese Umstände gekennzeichnet sind, zu der Anzahl der wahren und falschen Zeugenaussagen, die durch die angegebenen Umstände bestimmt sind".

Soll das heißen, daß man die relative Häufigkeit der "wahren" Zeugenaussagen in der Menge der "wahren" und "falschen" Zeugenaussagen in "allen" dem vorliegenden Fall "ähnlichen" Fällen bestimmen soll? Dann fragt man sich, wenn das überhaupt stimmt, wer soll imstande sein, die Anzahl "aller wahren und aller falschen" Aussagen zahlenmäßig zu bestimmen, geschweige denn, diese Aussagen auf ihre Wahrheit hin empirisch zu überprüfen²⁾. Und was heißt hier "ähnlich"??

Auch dann, wenn das möglich wäre, was hat die so festgestellte relative Häufigkeit mit der konkreten Zeugenaussage im vorliegenden Fall zu tun? Denn über die Glaubwürdigkeit eines Zeugen entscheidet, wie bekannt, nicht die Tatsache, daß andere Zeugen in soundso vielen früheren "ähnlichen" Fällen mit der Wirklichkeit übereinstimmen - und in soundso vielen falsche Aussagen gemacht haben, sondern

1) Glaubwürdigkeit einer Zeugenaussage (= Beweiswert dieser Zeugenaussage) wird definiert (vgl. die vorhergehende Seite) als die Wahrscheinlichkeit der Wahrheit dieses Aussage; Schreiber, S.17, 21.

2) vgl. S.103, Fn.6, dieser Arbeit.

3) bzw. in Schreibers Ausdrucksweise "die durch diese Umstände gekennzeichnet sind", "die durch die angegebenen Umstände bestimmt sind" (vgl. das Zitat).

in den allermeisten Fällen einzig und allein der persönliche Eindruck des Richters von diesem konkreten Zeugen; diesem Eindruck liegen u.a. seine Menschenkenntnisse und auch gewisse, meistens verschwommene induktive Regularitäten (etwa psychologischer Art) zugrunde, keineswegs aber die von Schreiber gemeinten Kenntnisse statistischer Art¹⁾

Analog der Zeugenglaubwürdigkeit wird der Beweiswert aller anderen Beweismittel als die Wahrscheinlichkeit der Wahrheit der Aussage definiert.

Schauen wir uns jetzt die von Schreiber vorgeschlagenen Regeln der Beweismittelwürdigung an.

1. Mehrere notwendige Beweismittel:

"Hier soll das Verfahren angegeben werden, das erforderlich ist, wenn mehrere Beweismittel notwendig sind, um den Beweis zu erbringen. Man kann es auch so ausdrücken, daß dies dann der Fall ist, wenn die Beweisthemen, die die Beweismittel betreffen, sich unterscheiden und kumulativ²⁾ zum Beweis der Behauptung erforderlich sind"³⁾

Die Formel lautet für n Beweismittel:

$$b = b_1 b_2 \dots b_n \quad \text{mit } b_i = \frac{w_i}{w_i + f_i},$$

d.h. b_i ist das Verhältnis der wahren Aussagen zu der Anzahl der wahren und falschen Aussagen für ein bestimmtes Beweismittel⁴⁾

(Diese Formel hat Ähnlichkeit mit dem Multiplikationsprinzip, vorausgesetzt, man nimmt die Unabhängigkeit der einzelnen Beweismittel an, was Schreiber übrigens nicht tut).

Die obige Formel wird von Schreiber wie folgt illustriert (S.25):

"Das Problem soll an einem Beispiel verdeutlicht werden. Läßt sich die Frage, ob eine bestimmte Verletzung von der Handlung eines Täters herrührt, nur mit der Wahrscheinlichkeit b_1 feststellen, und läßt sich die Frage, ob ein Schaden der sich aus der Verletzung ergab, nur mit der Wahrscheinlichkeit b_2 feststellen, so hängt die Wahrscheinlichkeit, daß der Schaden durch die Handlung des Täters verursacht wurde, von den Wahrscheinlichkeiten b_1 und b_2 ab".

(Diese wird dann anhand der Formel als $b = b_1 b_2$ berechnet).

1) Die Wahrscheinlichkeit der Wahrheit der Zeugenaussage als relative Häufigkeit, die für vergangene Fälle festgestellt ist, läßt sich auch in formaler Hinsicht nicht auf den neuen Fall übertragen; denn in diesem Fall ist die Bezugsmenge aller möglichen Fälle um ein neues Element vergrößert; für diese ist aber die relative Häufigkeit noch nicht festgestellt. Vgl. S. 68f dieser Arbeit, wo ich mich mit dieser Problematik ausführlicher befaße.

2) bei Schreiber: "kumulativ".

3) Vgl. Schreiber, S.25.

4) Vgl. ibid, S.25ff.

Wie wir aus dem obigen Zitat ersehen können, identifiziert Schreiber (unzulässigerweise) die Wahrscheinlichkeit für die Wahrheit eines Beweismittels mit der Wahrscheinlichkeit des Sachverhalts, den dieses Beweismittel bestätigt (d.h. mit der bedingten Wahrscheinlichkeit des Sachverhalts aufgrund des Beweismittels)¹⁾

Kürzen wir "das Vorliegen der Handlung" mit H, "das Vorliegen der Verletzung" mit V und "das Vorliegen des Schadens" mit S ab, so erhalten wir:

$$P(V/H) = b_1, P(S/V) = b_2.$$

$b_1 \cdot b_2$ ($= P(S/V) \cdot P(V/H)$) ist dann die bedingte Wahrscheinlichkeit des Sachverhalts (= "Schadens") aufgrund des Beweismittels (= "Verletzung") multipliziert mit der (bedingten) Wahrscheinlichkeit des Beweismittels aufgrund einer weiteren Information ("Handlung").

Nach Schreiber muß $b_1 \cdot b_2 = P(S/H)$, d.h. $P(S/V) \cdot P(V/H) = P(S/H)$, gelten. Dies stimmt jedoch, wie der nachfolgende Versuchsfall zeigt (der analog dem obigen Beispiel konstruiert ist) nicht immer.

Angenommen, in einer Urne befinden sich 10 durchnummerierte Kugeln. Einige davon sind schwarz und einige sind mit einem Punkt markiert, vgl. die Darstellung unten:

1	2	3	4	5	6	7	8	9	10
	x		x	x			x		
	.				.		.		

(mit x sind schwarze und mit . sind markierte Kugeln gekennzeichnet)

Die Wahrscheinlichkeit, eine schwarze Kugel zu ziehen, wenn bekannt ist, daß sie eine gerade Nummer trägt, ist $3/5$; die Wahrscheinlichkeit, eine markierte Kugel zu ziehen, vorausgesetzt, sie ist schwarz, ist $1/2$. Die Wahrscheinlichkeit, eine markierte Kugel zu ziehen, unter der Voraussetzung, daß sie eine gerade Nummer trägt, $3/5$ und nicht, folgt man Schreiber, $3/5 \cdot 1/2 = 3/10$. Die Wahrscheinlichkeit von $3/10$ wäre nur dann möglich, wenn die Menge der markierten Kugeln im obigen Fall eine Teilmenge der Menge der schwarzen Kugeln und die Menge der schwarzen eine Teilmenge derjenigen mit den geraden Nummern wäre; dies wird aber weder hier noch in Schreibers Beispiel explizit vorausgesetzt; in Schreibers Beispiel entspricht dem Fehlen einer solchen Voraussetzung die Annahme, daß eine (strafbare) Handlung zwar zu einem Schaden, aber nicht dabei unbedingt zu einer Verletzung führen muß.

Die richtige Formel für das von Schreiber angeführte Beispiel muß somit lauten:

$$P(S/H) = P(S/V \wedge H) \cdot P(V/H) + P(S/\neg V \wedge H) \cdot P(\neg V/H)^2$$

1) Unter Beweismitteln versteht man neben Zeugen- und Sachverständigenaussagen auch Urkunden, Augenschein und Indizien (vgl. E.Döring, S.19); d.h. im normalen Sprachgebrauch sind es Daten bzw. Informationen, aufgrund derer man den Sachverhalt beurteilt.

2) Vgl. die Bayes-Formel (eine Erweiterung der Bayes-Formel), S.34 dieser Arbeit. " $P(\neg V/H)$ " in der obigen Formel ist zu lesen als: "Die Wahrscheinlichkeit dafür, daß keine Verletzung aufgrund der strafbaren Handlung vorliegt".

Abgesehen von der obigen Schwierigkeit muß man genau unterscheiden, für welche Fälle die obige Formel ($b = b_1 b_2 \dots b_n$) anwendbar ist und für welche nicht. So ist sie dann anwendbar, wenn wir die Wahrscheinlichkeit für das gemeinsame Vorliegen mehrerer Beweismittel bzw. mehrerer Sachverhalte (ihre Unabhängigkeit vorausgesetzt) ermitteln wollen. Sie ist nicht anwendbar, wenn wir die Wahrscheinlichkeit eines Sachverhalts aufgrund mehrerer Beweismittel bestimmen wollen¹⁾ Schreiber meint aber offensichtlich auch diesen Fall, wenn er schreibt (S.27):

"Die hier angegebene Formel ist nicht nur erforderlich, wenn mehrere Zeugen Aussagen machen, die nur zusammengekommen die Behauptung genau bestätigen. Auch in Gutachten, die sich auf verschiedene Fragen beziehen, aber beide zur Tatsachenfeststellung erforderlich sind, ist eine Methode erforderlich, eine Gesamtaussage über die Wahrscheinlichkeit des Ereignisses zu bilden"²⁾

2. Mehrere hinreichende Beweismittel:

Es handelt sich um den Fall, daß "mehrere Beweismittel zu einem Beweissthema zur Verfügung stehen, die bereits für sich allein - bei ausreichendem Beweiswert - zur Feststellung der Wahrheit der Behauptung ausreichen würden"³⁾

Für den Fall mehrerer hinreichend bestätigender Beweismittel gibt Schreiber die folgende Formel an (S.30f):

$$b = \frac{b_1 \cdot b_2}{b_1 b_2 + (1-b_1)(1-b_2)} \quad \text{mit } b_i = \frac{w_i}{w_i + f_i}$$

$$b = \frac{w_1 \cdot w_2}{w_1 w_2 + f_1 f_2} \quad (w_i, f_i \text{ wie früher})$$

Für n Beweismittel erhalten wir dementsprechend:

$$b = \frac{b_1 \cdot b_2 \dots b_n}{b_1 b_2 \dots b_n + (1-b_1)(1-b_2) \dots (1-b_n)}$$

(Voraussetzung ist allerdings, daß die n Beweismittel voneinander unabhängig sind).

1) Eine solche Bestimmung kann nur mittels der Formel für die bedingte Wahrscheinlichkeit bzw. mit dem Bayes-Theorem erfolgen.

2) Haben wir nämlich zwei Beweismittel (z.B. Zeugenaussagen) D_1 und D_2 mit $b_1 = b_2 = 0.5$ (bzw. $P(D_1) = P(D_2) = 0.5$) - wir nehmen an, daß sie voneinander unabhängig sind -, so ergibt Schreibers Formel $b = 0.25$, d.h. die Wahrscheinlichkeit für das gemeinsame Vorliegen dieser Beweismittel, nicht aber die Wahrscheinlichkeit für den Sachverhalt aufgrund dieser Beweismittel; auf gar keinen Fall entspricht Schreibers Formel seiner Intention, eine Methode anzugeben, mit der man den Schluß, daß mehrere Beweismittel "nur zusammen die Behauptung genau bestätigen" (vgl. das obige Zitat), rechtfertigen kann.

3) Schreiber S.29.

Es ist interessant zu wissen, wie Schreiber zu dieser Formel gekommen ist¹⁾

Auf S. 30f ist die folgende Äußerung zu finden:

"Die Beweissituation, in der zwei Beweismittel eine Behauptung bestätigen, ist dadurch gekennzeichnet, daß beide Aussagen der Beweismittel nur zusammen wahr oder beide zusammen falsch sein können. Denn sie beziehen sich auf die gleiche Behauptung. Diese kann aber nicht bei gleichen Aussagen durch das eine Beweismittel bestätigt und gleichzeitig durch das andere widerlegt werden. Das heißt, die Kombinationen w_1f_2 und w_2f_1 sind nicht möglich. Ihre Anzahl ist demnach gleich null.

Bei $w_1f_2 = 0$ und $w_2f_1 = 0$

erhalten wir
$$b = \frac{w_1w_2}{w_1w_2 + f_1f_2}.$$

Schreiber behauptet also, daß der Fall unmöglich ist, daß ein Beweismittel wahr und ein anderes falsch ist, während sie ein und dieselbe Behauptung bestätigen. Diese Auffassung können wir jedoch inhaltlich leicht widerlegen, indem wir ein Gegenbeispiel anführen:

Angenommen, wir haben eine Zeugenbeschreibung des Täters, die einwandfrei auf den Angeklagten paßt, und eine weitere Zeugenaussage, in welcher vom Zeugen behauptet wird, daß der Angeklagte in seiner Gegenwart die Tat angekündigt hat. Wie wir sehen, bestätigen die beiden Zeugenaussagen die Annahme, daß der Angeklagte der Täter ist; sie können aber trotzdem in ihrem Wahrheitswert sich voneinander unterscheiden: So kann etwa der erste Zeuge den Täter wahrheitsgetreu beschrieben und der zweite Zeuge gelogen haben; die erste Aussage ist somit wahr, die zweite falsch²⁾.

1) Schreiber lehnt sich dabei offensichtlich an die Boolesche Formel an (vgl. S. 85 dieser Arbeit). Zur Kritik an dieser Formel (und an Schreiber) vgl. auch Cohen S. 95ff.

2) Schreiber gibt noch weitere Formeln an:

Für mehrere hinreichend widerlegende Beweismittel:

$$b = \frac{(1-b_1)(1-b_2)\dots(1-b_n)}{b_1b_2\dots b_n + (1-b_1)(1-b_2)\dots(1-b_n)} \quad \text{mit } b_i = \frac{w_i}{w_i + f_i}, \quad i = 1, \dots, n,$$

als einzelne widerlegende Beweismittel;

für $n = 2$ lautet die Formel:

$$b = \frac{f_1f_2}{w_1w_2 + f_1f_2};$$

Für mehrere sich widersprechende Beweismittel ist die Formel viel komplizierter; wir geben sie nur für den Fall an, wo zwei sich widersprechende Beweismittel vorliegen:

$$b = \frac{b_b(1-b_w)}{b_b(1-b_w) + (1-b_b)b_w}; \quad b_b \text{ ist das bestätigende und } b_w \text{ das widersprechende Beweismittel}$$

Da diese Formeln auf die Formel für hinreichend bestätigende Beweismittel zurückführbar sind (vgl. Schreiber, S. 35, 38), gilt auch hier die von mir vorgetragene Kritik.

Allerdings läßt sich, wie Koch&Rüßmann vorbringen¹⁾, Schreibers Formel retten, wenn man sie auf direkte Beobachtungen bezieht bzw. wenn Zeugen wahrheitsgetreu über ihre Beobachtungen berichten. Da wir es in der Gerichtspraxis weder mit direkten Beobachtungen noch mit Zeugenaussagen als wahrheitsgetreuen Abbildungen der Wirklichkeit zu tun haben, scheint ein solcher Rettungsversuch wenig zu bringen.

Für Schreibers Formel gilt die gleiche Kritik wie für die Boolesche Formel²⁾, nämlich, daß sie die Wahrscheinlichkeit der Hypothese nicht berücksichtigt bzw., wenn man sie bayesianistisch interpretiert, nur für die Apriori-Wahrscheinlichkeit 0.5 korrekt ist; in diesem Fall stimmt sie mit der Bayes-Formel überein, wobei dies wiederum nur eingeschränkt gilt: Denn die Bayes-Formel verlangt keineswegs, daß die Wahrscheinlichkeit von Beweismitteln unter der Falschheit der Hypothese gleich $1-b$ ist (vgl. Schreibers Formel für hinreichende Beweismittel, wenn b bayesianistisch als Wahrscheinlichkeit der Beweismittel unter der Voraussetzung der Wahrheit der Hypothese interpretiert wird).

Man kann fragen, ob es überhaupt zulässig ist, die W. für die Aussage aufgrund der Falschheit der Hypothese mit der W. der nicht korrekten Aussage aufgrund der Hypothese zu identifizieren (wie das bei den bayesianistisch interpretierten Formeln von Boole und Schreiber der Fall ist). In dem auf S.84 dieser Arbeit geschilderten Experiment von Inbau geht es um die Beurteilung der Korrektheit bzw. Nichtkorrektheit von Zeugenaussagen relativ zu allen (wahren oder falschen) Hypothesen, d.h. Inbau will die Wahrscheinlichkeiten $P(A_k/H_e \vee H_{ue})$ und $P(A_{nk}/H_e \vee H_{ue})$ in einer Anzahl von Fällen ermitteln (" A_k " ist "korrekte Zeugenaussage", " A_{nk} " "nicht korrekte Zeugenaussage", " H_e " "echte Unterschrift", " H_{ue} " "unechte Unterschrift"). Diese beiden Wahrscheinlichkeiten gehen also als P und $1-P$ in die Boolesche Formel ein. In dem von Ball (S.85 dieser Arbeit) beschriebenen Fall geht es aber, da die Hypothese über die Echtheit der Unterschrift beurteilt wird, dagegen um die W. einer (korrekten) Zeugenaussage relativ zu einer wahren Hypothese, d.h. $P(A_e/H_e)$ und um die W. einer (nicht mehr korrekten) Zeugenaussage relativ zu einer falschen Hypothese, d.h. $P(A_e/H_{ue})$, falls die Zeugenaussage die Echtheit der Unterschrift bestätigt, bzw. um $P(A_{ue}/H_e)$ und $P(A_{ue}/H_{ue})$, falls die Zeugenaussage gegen die Echtheit spricht (" A_e " steht für: "Die Zeugenaussage lautet: 'die Unterschrift ist echt'", " A_{ue} ": "Die Zeugenaussage lautet: 'die Unterschrift ist unecht'").

1) Vgl. Koch&Rüßmann, Juristische Begründungslehre (1982), S.311: "Beruhen die Beobachtungsberichte auf voneinander unabhängigen Beobachtungen, so läßt sich der Gesamtwert für die fragliche Sachverhaltsannahme relativ zu den Einzelwerten zuverlässig mit der Formel für hinreichende Beweismittel errechnen. Diese Interpretation reserviert der Arbeit Schreibers zwar einen ungeheuer schmalen Anwendungsbereich; sie macht allerdings auch Schreibers Ergebnisse unangreifbar".
2) Vgl. S.86, Fn.1 dieser Arbeit.

D.h. es wird bei der Einsetzung in die Boolesche Formel $P(A_k/H_e \vee H_{ue})$ mit $P(A_e/H_e)$, falls die Zeugenaussage die Echtheit der Unterschrift bestätigt (bzw. mit $P(A_{ue}/H_{ue})$, falls die Zeugenaussage die Unechtheit der Unterschrift bestätigt) und $P(A_{nk}/H_e \vee H_{ue})$ mit $P(A_e/H_{ue})$ (bzw. mit $P(A_{ue}/H_e)$) unzulässigerweise identifiziert. Addieren sich die beiden Wahrscheinlichkeiten $P(A_k/H_e \vee H_{ue})$ und $P(A_{nk}/H_e \vee H_{ue})$ zu 1, so ist das bei den Wahrscheinlichkeiten $P(A_e/H_e)$ und $P(A_e/H_{ue})$ bzw. $P(A_{ue}/H_e)$ und $P(A_{ue}/H_{ue})$ nicht der Fall.

Um die Wahrscheinlichkeiten $P(A_e/H_e)$ und $P(A_e/H_{ue})$ bzw. $P(A_{ue}/H_e)$ und $P(A_{ue}/H_{ue})$ feststellen zu können, muß das Experiment von Inbau anders lauten: 1.(die Zeugenaussage bestätigt die Echtheit der vorliegenden Unterschrift): Man ermittelt die W. der korrekten Aussage, wenn die Hypothese über die Echtheit der Unterschrift wahr ist, d.h. $P(A_e/H)$ und die W. der nicht korrekten Aussage, wenn die Hypothese über die Echtheit der Unterschrift falsch ist, d.h. $P(A_e/\neg H)$, mit $H \hat{=} H_e$ und $\neg H \hat{=} H_{ue}$; 2.(die Zeugenaussage spricht gegen die Echtheit der vorliegenden Unterschrift): Man ermittelt die W. einer korrekten Aussage (über die Nichtechtheit der vorliegenden Unterschrift), falls die Hypothese über die Echtheit der Unterschrift falsch ist, d.h. $P(A_{ue}/\neg H)$ und die W. einer nicht korrekten Aussage, falls die Hypothese über die Echtheit wahr ist, d.h. $P(A_{ue}/H)$. Diese Wahrscheinlichkeiten sollen dann in die Boolesche Formel eingesetzt werden.

Es ist klar, daß die beiden Wahrscheinlichkeiten $P(A_e/H)$ und $P(A_e/\neg H)$ bzw. $P(A_{ue}/H)$ und $P(A_{ue}/\neg H)$ sich nicht zu 1 zu addieren brauchen. So gesehen läßt sich die Boolesche Formel, aber auch (im Widerspruch zur Behauptung von Koch&Rüßmann) Schreibers Formel (auch dann, wenn es um direkte Beobachtungen geht) angreifen.

Ein anderer Grund, die Boolesche Formel abzulehnen, ist bereits angesprochen worden, nämlich, daß die Glaubwürdigkeit der zu beurteilenden Hypothese nicht berücksichtigt wird. Ist etwa eine solche sehr gering, so führt die Boolesche Formel zur Überschätzung von Beweismitteln, ist diese dagegen hoch, so kann es passieren, daß die Boolesche Formel trotzdem zur Verwerfung der Hypothese auch dann führen würde, wenn die Beweismittel für sich genommen zwar einen "nicht genug hohen", für die Hypothese allerdings ausreichenden "Beweiswert" haben.

Resümee:

Den Versuch Schreibers, eine Theorie der Beweiswürdigung aufgrund der mathematischen Wahrscheinlichkeitstheorie aufzustellen¹⁾, können wir als mißlungen ansehen. Denn die Regeln, die Schreiber vorschlägt, sind inhaltlich nicht adäquat oder sie beruhen auf für die Rechtspraxis nicht haltbaren Voraussetzungen.

So gibt die Formel für mehrere notwendige Beweismittel lediglich die Wahrscheinlichkeit für das gemeinsame Vorliegen mehrerer Beweismittel bzw. mehrerer Sachverhalte²⁾ (ihre Unabhängigkeit voneinander

1) Vgl. Schreiber (S.24): "Für das Maß der Glaubwürdigkeit empfiehlt sich der Maßstab, den die mathematische Wahrscheinlichkeitstheorie zugrunde legt. Die Werte umfassen den Bereich von 0 bis 1....".

2) Wie die Bezeichnung "Gesamtwahrscheinlichkeit mehrerer notwendiger Beweismittel" besagt, bezieht sich Schreibers Formel auf Beweismittel (also nicht auf Sachverhalte); dies steht allerdings in Dis-

vorausgesetzt), keineswegs jedoch die von Schreiber offensichtlich gewollte Wahrscheinlichkeit einer Behauptung aufgrund dieser Beweismittel an. Denn diese wird nicht durch das Multiplizieren der Wahrscheinlichkeiten für einzelne Beweismittel, sondern aufgrund der Formel für die bedingte Wahrscheinlichkeit bestimmt.

Während man die Formel für notwendige Beweismittel in den Fällen, wo die Unabhängigkeit der Beweismittel garantiert ist, als eine Art Multiplikationsprinzip akzeptieren kann, so ist die Formel für mehrere hinreichende bestätigende Beweismittel völlig wertlos. Denn sie (und mit ihr alle anderen auf S.108, Fn.2, angeführten Formeln) geht nicht nur von der inhaltlich falschen Voraussetzung aus, daß Beweismittel nur dann eine Behauptung bestätigen können, wenn sie den gleichen Wahrheitswert aufweisen, sondern berücksichtigt in keiner Weise die Wahrscheinlichkeit der zu beurteilenden Hypothese und außerdem - wenn man die Formel bayesianistisch interpretiert - die Tatsache, daß die Wahrscheinlichkeiten von Beweismitteln zum einen relativ zur Hypothese und zum anderen relativ zur Negation der Hypothese gewöhnlich beliebige, voneinander unabhängige Werte haben (d.h. die beiden addieren sich nicht, wie die Formeln von Schreiber und Boole es verlangen, zu 1); darüber hinaus läßt sie sich ohne weiteres durch die korrekte Bayes-Formel ersetzen.

krepanz zu dem von Schreiber angegebenen und von mir auf S.105f dieser Arbeit kritisierten Beispiel, wo die Formel für "mehrere notwendige Beweismittel" sowohl auf Beweismittel wie auch auf einen Sachverhalt angewandt wird.

1.6. Schilderung des Rechtsfalls 'People v. Collins'

Bis jetzt haben wir einige theoretische Versuche diskutiert, die Wahrscheinlichkeitstheorie und die Statistik für die Rechtspraxis nutzbar zu machen. Wie wir gesehen haben, waren diese Versuche nicht ohne Fehler und falsche Ansätze. Später folgte, insbesondere als man anfang, auch konkrete Fälle in der Gerichtspraxis mit wahrscheinlichkeitstheoretischen Methoden zu lösen, eine Reihe von Aufsätzen¹⁾, die vor allem den leichtfertigen und unreflektierten Gebrauch der statistischen Methoden für die Lösung juristischer Fälle kritisierten. Am gründlichsten ist dabei Cullison vorgegangen.. Bevor wir auf seinen Aufsatz näher eingehen, wollen wir den Rechtsfall schildern, an dem sich eine beinahe 10 Jahre lang währende kritische Auseinandersetzung mit der praktischen Anwendung der Wahrscheinlichkeitstheorie in der Rechtspraxis entzündete.

People v. Collins²⁾

In diesem Rechtsfall verwendete die Anklagevertretung eine wahrscheinlichkeitstheoretische Argumentation, um die Identität des angeklagten Paares mit dem Täterpaar zu beweisen. Diese Argumentation war, wie die spätere Kritik zeigt, v.a. aus drei Gründen fehlerhaft: erstens, man nahm unbegründeterweise die Unabhängigkeit der vorliegenden Befunde an, zweitens, man stützte die statistischen Daten, die der probabilistischen Kalkulation zugrunde gelegt wurden, nicht empirisch und man wählte drittens den falschen Ansatz bei der Wahrscheinlichkeitsbestimmung der Identität.

Der Fall wird folgendermaßen geschildert³⁾

Janet und Malcolm Collins wurden von einem Gericht in Los Angeles eines Raubüberfalls für schuldig befunden. Malcolm legte eine Berufungsklage bei dem Supreme Court of California ein und der Schuldspruch wurde revidiert. Das Gericht beschrieb den Raubüberfall so:

Am 18. Juni 1964, gegen 11:30 a.m. war Mrs. Juanita Brooks nach ihren Einkäufen auf dem Wege nach Hause; sie ging eine Allee im San Pedro Bezirk der Innenstadt von Los Angeles entlang. Hinter sich zog Mrs. Brooks einen Einkaufswagen mit Lebensmitteln her, auf dem ganz oben ihre Geldbörse lag. Als sie sich bückte, um einen leeren Karton aufzuheben, wurde sie plötzlich von einer Person zum Boden gestoßen, die sie weder kommen gesehen noch gehört hatte. Durch den Fall war sie wie betäubt. Es gelang ihr jedoch aufzuschauen; dabei sah sie, wie eine junge Frau davonlief. Nach der Beschreibung,

1) Vgl. W.B. Stoebe, *Relevancy and the Theory of Probability*, in: 51 Iowa Law Rev. (1966), S. 849-861; T.H. Liddle, *Mathematical and Statistical Probability as a Test of Circumstantial Evidence*, in: 19 Case Western Law Rev. (1968), S. 254-78; P. Sigal, *Judicial Use, Misuse, and Abuse of Statistical Evidence*, in: 47 Journal of Urban Law (1968), S. 165-90; A.D. Cullison, *Identification by Probabilities and Trial by Arithmetic (A Lesson for Beginners in How to be Wrong with Greater Precision)*, in: 6 Houston Law Rev. (1969), S. 471-518. 2) 66 Cal. Rptr. 497 (1968).

3) Die Schilderung dieses Falls wurde im wesentlichen aus dem Aufsatz von Fairley und Mosteller, *A Conversation About Collins*, in: 41 Univ. of Chicago Law Rev. (1974), S. 242-53, übernommen.

die Mrs. Brooks gab, wog diese ca. 145 Pfund, trug "etwas schwarzes" und war "zwischen dunkelblond und hellblond", allerdings heller als die Angeklagte Janet Collins, als sie vor Gericht erschien. Gleich nach dem Überfall vermißte Mrs. Brooks ihre Geldbörse, die \$35 bis \$40 enthielt.

Zur gleichen Zeit befand sich John Bass, der am Ende der Allee wohnte, beim Gießen des Rasens vor seinem Haus. Er wurde auf das Geschrei, das aus der Allee kam, aufmerksam. Als er in jene Richtung schaute, sah er eine Frau, die aus der Allee hinauslief und in ein gelbes Auto einstieg, das ihm gegenüber auf der anderen Strassenseite geparkt war. Den Wagentyp konnte er jedoch nicht erkennen. Als das Auto ganz nah, im Abstand von ca 2 m, an ihm vorbei fuhr, sah er, daß es von einem männlichen Schwarzen mit Bart und Schnurrbart gesteuert wurde. Vor Gericht konnte Bass den Angeklagten als den Fahrer des gelben Autos identifizieren. Allerdings wurde ein Versuch unternommen, seine Identifizierung anzuzweifeln, als sich herausstellte, daß Bass bei der ersten Gegenüberstellung kurz nach dem Raubüberfall Schwierigkeiten hatte, Malcolm Collins, der zu dieser Zeit keinen Bart trug, als Täter zu identifizieren.

In seiner Aussage beschrieb Bass die Frau, die vom Tatort weglief, als vom kaukasischen Typ, etwas über 1.50 m groß, normal gebaut, mit dunkelblondem Haar in Pferdeschwanzfrisur, "genau so", wie sie Janet auf einem Polizeifoto hatte, das am 22. Juni 1964 aufgenommen wurde.

Vor Gericht führte der Anklagevertreter mit Hilfe eines Mathematikdozenten eine Tabelle ein, in welcher er hypothetischerweise folgende Wahrscheinlichkeiten für das Auftreten einzelner Merkmale angab:

Merkmale:	einzelne Wahrscheinlichkeiten:
A. Zum Teil gelbes Auto	1/10
B. Mann mit Schnurrbart	1/4
C. Mädchen mit Pferdeschwanz	1/10
D. Mädchen mit blondem Haar	1/3
E. Schwarzer Mann mit Bart	1/10
F. Schwarz-weißes Paar im Auto	1/1000 ¹⁾

Der Anklagevertreter multiplizierte diese einzelnen Wahrscheinlichkeiten und erreichte einen Wert von 1/12000000, der die Wahrscheinlichkeit repräsentieren sollte, "daß ein beliebiges Paar die gleichen Merkmale haben würde, wie das angeklagte Paar". Anschließend resümierte er, daß es extrem unwahrscheinlich sei - "ungefähr eins zu 12 Millionen"²⁾ - daß ein anderes als das angeklagte Paar all diese Merkmale hätte.

1) Wie wir sehen, unterscheidet sich diese Tabelle in zwei Hinsichten von den entsprechenden Stellen der vorangegangenen Beschreibung: Es wird ein zum Teil gelbes Auto angenommen (aus dem Text geht hervor, daß es ein gelbes Auto war) und die Haarfarbe des Mädchens wird weiter nicht präzisiert, obwohl gerade das wichtig wäre, da Mrs. Brooks der Auffassung ist, daß die Täterin helleres Haar hatte als Janet Collins.

2) Vgl. Stripinis, S.64: "Applying the product rule to his own factors the prosecutor arrived at a probability that there was but one chance in 12 million that any couple possessed the distinctive characteristics of the defendants. Accordingly, under this theory, it was to be inferred that there could be but one chance in 12 million that defendants were innocent and that another equally distinctive couple actually committed the robbery".

Auf eine Reihe von Fehlern bei der obigen Wahrscheinlichkeitsanalyse hat bereits Kingston hingewiesen, der diesen Fall erstmals untersucht hat¹⁾. Diese sind: 1) Die Ungenauigkeit hinsichtlich des Umfangs der Bezugsklasse ($\hat{=}$ Bevölkerungsgruppe), relativ zu welcher statistische Schätzungen gemacht werden bzw. die Verschiedenheit der jeweiligen Bezugsklassen, 2) die unbegründete Annahme der Unabhängigkeit zwischen einigen Merkmalen (wie etwa Schnurrbart und Bart), 3) die nicht korrekte Annahme, daß die verschwindend kleine Wahrscheinlichkeit für das gemeinsame Vorkommen aller genannten Merkmale (die durch das Multiplizieren einzelner Wahrscheinlichkeiten erreicht wird) ausreicht, das verdächtige Paar als das Täterpaar zu identifizieren. Nach Kingston muß man stattdessen ermitteln: erstens, die Wahrscheinlichkeit für das Auffinden eines weiteren Paares mit den gleichen Merkmalen wie bei dem angeklagten Paar in der fraglichen Bevölkerung und zweitens, die Irrtumswahrscheinlichkeit, wenn man das verdächtige Paar mit dem Täterpaar identifiziert. (Ist nämlich die Wahrscheinlichkeit dafür, daß ein anderes Paar mit solchen Merkmalen existiert, hoch, dann kann man nicht mehr so sicher sein, daß das verdächtige Paar auch das richtige ist, auch dann, wenn die Wahrscheinlichkeit für das gemeinsame Auftreten aller Merkmale sehr gering ist²⁾).

Neben den angeführten statistischen Daten³⁾ wollen wir hier auch nichtstatistische Daten, die für den Collins-Fall relevant sind, angeben. Anzumerken ist, daß in allen Versuchen, diesen Fall wahrscheinlichkeitstheoretisch zu lösen, nur statistische Daten berücksichtigt wurden. Das verwundert allerdings nicht, da man stets den statistischen Wahrscheinlichkeitsbegriff den Beurteilungen zugrunde gelegt hat. Will man jedoch neben statistischen Daten solche nicht statistischer Art berücksichtigen, so muß man den subjektiven Wahrscheinlichkeitsbegriff wählen.

Nicht-statistische Daten (aus der vorangegangenen Schilderung des Falls):

Die Zeugenaussage von Mrs. Brooks: J. Brooks beschrieb die Täterin als eine junge Frau, die "etwas schwarzes" trug, ca 145 Pfund wog und "zwischen dunkelblond und hellblond", jedoch heller als Janet Collins, war.

Die Zeugenaussage von John Bass: J. Bass beschrieb die Frau, die aus der Allee, aus der er kurz zuvor Schreie vernommen hatte, hinauslief und in ein gelbes Auto einstieg als vom kaukasischen Typ, etwas über 150 m groß, normal gebaut, mit dunkelblondem Haar⁴⁾.

Zusätzliche nicht-statistische Daten (aus Cullisons Schilderung dieses Falls⁵⁾):

Die Aussagen anderer als der beiden oben erwähnten Zeugen:

Zeugen haben ausgesagt, daß am Tag Janet Collins eine Pferdeschwanzfrisur trug, daß ihr schwarzer Ehemann, der Angeklagte Malcolm Collins, einen Bart und einen

1) C.R. Kingston, Probability and Legal Proceedings, in: 57 Journal of Criminal Law, Criminology and Police Science (1966), S. 93-98. Kingston bietet außerdem bei der Berechnung der Wahrscheinlichkeit für das gemeinsame Auftreten dieser Merkmale eine Lösung an, die die vom Gericht gemachten Fehler vermeidet.

2) Wie man die Irrtumswahrscheinlichkeit bestimmen kann, wurde von Kingston in seinem Aufsatz "Applications of Probability Theory in Criminalistics" gezeigt, vgl. S 92ff dieser Arbeit.

3) Unter den statistischen Daten wollen wir hier die Daten verstehen, die über eine Gesamtheit von Fällen, nicht aber über einen einzelnen Fall, aussagen.

4) Die anderen Daten über die Täterin und Bass' Beschreibung des Täters wurden bereits bei der statistischen Beurteilung berücksichtigt.

5) Vgl. Cullison (1969), S. 475, auch Fn. 10, 11 (ibid).

Schnurrbart hatte und einen gelben Lincoln mit weißem Dach fuhr. Außerdem wurde bestätigt, daß Janet Collins den Arbeitsplatz ungefähr zur Zeit des Überfalls verließ.

Die Zeugen waren sich nicht einig, ob das Auto gelb mit weißem Dach oder nur gelb war.

Die Augenzeugen des Überfalls waren sich darüber uneinig, ob Janets Haar dunkler war als das der Täterin. Man konnte aber nachweisen, daß Janet versucht hatte, ihre Haarfarbe nach dem Tag des Raubüberfalls zu ändern. Auch Malcolm hatte sich rasiert, obwohl er gewöhnlich einen Bart trug. Es war jedoch nicht klar, ob er es vor dem Raubüberfall oder erst danach getan hat.

Als die Polizisten zu Malcolms Haus kamen, um ihn zu verhaften, lief er nach ihrer Aussage durch die Hintertür weg und versteckte sich auf der Toilette des Nachbarn. Am Tag nach dem Überfall zahlte Malcolm eine Geldstrafe von ungefähr dem gleichen Betrag, der gestohlen wurde; zur Quittung, die bei ihm gefunden wurde, erklärte er, daß er das Geld in einem Glücksspiel gewonnen habe.

Janet Collins machte während des Verhörs Andeutungen, daß sie gestehen würde, wenn man ihren Mann freiließe.

Die Angeklagten ihrerseits beteuerten, bei einem weit vom Tatort entfernt wohnenden Freund zu Besuch gewesen zu sein. Der Freund bestätigte das, konnte allerdings nicht genau das Datum des Besuchs nennen. Ansonsten bestritten die Angeklagten, die Tat begangen zu haben.

Wie wir sehen, enthielt der Collins-Fall neben statistischen Daten genug an nichtstatistischer Information, um das angeklagte Paar als das Täterpaar zu verurteilen. Als stärkstes Argument gegen Collins führte der Anklagevertreter das statistische Datum an. Ironischerweise konnte der Supreme Court durch die Zahlenmanipulation (die Annahme einer unrealistisch hohen Zahl von $12 \cdot 10^6$ Paaren, die sich zur Tatzeit in der Nähe des Tatorts aufhielten) nachweisen, daß es eine ziemlich hohe Wahrscheinlichkeit von 0.42 gab, unter diesen 12000000 Paaren mindestens noch ein weiteres mit den gleichen Merkmalen zu finden. Obwohl es angesichts des Vorhandenseins einer ganzen Menge an "erdrückendem" Beweismaterial gegen Collins dem gesunden Menschenverstand widerspricht, wurden sie freigesprochen.

1.7. Cullison¹⁾

Cullison kritisiert die herkömmliche juristische Praxis, weil sie die Unsicherheit nicht berücksichtigt und weil sie versucht, allein mit der zwei-wertigen Logik auszukommen²⁾ Er hält die Wahrscheinlichkeitstheorie und Statistik für eine viel geeignetere Technik, mit Irrtümern und Unsicherheiten in der Rechtspraxis fertig zu werden³⁾ Außerdem ist er der Ansicht, daß ein mit der Wahrscheinlichkeitstheorie nicht ganz vertrautes Gericht diese leicht mißverstehen und mißbrauchen kann⁴⁾ Der größte Teil seines Aufsatzes ist daher der Einführung in die elementare Wahrscheinlichkeitsrechnung, der korrekten Bestimmung der Wahrscheinlichkeit für die Identität des verdächtigen Paares mit dem Täterpaar im Collins-Fall und der Kritik an den von der Rechtsprechung bei der wahrscheinlichkeitstheoretischen Beurteilung juristischer Fälle bis dahin begangenen Fehlern gewidmet.

Wie Kingston kritisiert auch Cullison das Vorgehen des Anklagevertreters im Collins-Fall, welcher die Wahrscheinlichkeit für das gemeinsame Auftreten der gleichen Merkmale bei einem zufällig gefundenen Paar (die aufgrund des Multiplikationsprinzips berechnet wurde) dafür benutzt, um zu zeigen, daß das angeklagte Paar auch das Täterpaar ist (so wird die Wahrscheinlichkeit für das Komplement, die gleich $11999999/12000000$ ist, mit der Wahrscheinlichkeit für die Täterschaft des angeklagten Paares und die Wahrscheinlichkeit $1/12000000$ mit einer solchen für seine Unschuld identifiziert).

Im folgenden soll das Vorgehen Cullisons bei der Bestimmung der Wahrscheinlichkeit für die Täterschaft kurz geschildert werden.

Wir benutzen folgende Abkürzungen:

p: die Wahrscheinlichkeit, daß ein zufällig gewähltes Paar die sechs Merkmale aufweist, die das Täter- und das angeklagte Paar besitzen (p ist im Collins-Fall = $1/12000000$);

1) A.D.Cullison, Identification by Probabilities and Trial by Arithmetic (A Lesson for Beginners in How to be Wrong with Greater Precision), in: 6 Houston Law Rev. (1969), S.471-518.

2) ibid, S.471f.

3) ibid, S.473.

4) Vgl. dazu auch die bereits dargelegte Kritik Kingstons an dem fehlerhaften Gebrauch der Wahrscheinlichkeitstheorie seitens des Gerichts, S.114 dieser Arbeit.

- n : die Anzahl der Paare aus einer bestimmten Population (bzw. Stichprobe), die auf das verdächtige Paar hin überprüft wird;
 k : die Anzahl der Paare in der Stichprobe, die die gleichen Merkmale aufweisen wie das Täterpaar.

Wir wissen zwar, daß nach Voraussetzung jedes der n Paare mit der Wahrscheinlichkeit p die fraglichen Merkmale hat, dagegen wissen wir nicht, wieviele unter n Paaren tatsächlich diese Merkmale besitzen (d.h. die Anzahl k ist uns völlig unbekannt).

Für den Fall, daß uns p und n bekannt sind, k dagegen nicht, bietet die Statistik eine Berechnungsformel – die sog. Binomialformel – an.

Nach einer weiteren Voraussetzung wissen wir, daß wenigstens ein Paar unter n Paaren die gleichen Merkmale wie das Täterpaar hat – dies sind nämlich die Angeklagten. Wir müssen somit die Wahrscheinlichkeit berechnen, daß dieses Paar mit dem Täterpaar identisch ist unter der Bedingung, daß es unter n Paaren mindestens ein Paar mit den in Frage kommenden Merkmalen gibt.

$P(b \equiv a / \geq 1)$ sei die bedingte Wahrscheinlichkeit, daß das verdächtige Paar auch das Täterpaar ist unter der Bedingung, daß es wenigstens ein solches (verdächtiges) Paar gibt.

Nach der Formel für die bedingte Wahrscheinlichkeit ist diese gleich:

$$P(b \equiv a / \geq 1) = \frac{P(b \equiv a \wedge \geq 1)}{P(\geq 1)} = \frac{P(b \equiv a)}{P(\geq 1)}$$

Den Zähler können wir wie folgt umformen:

$$\begin{aligned} P(b \equiv a) &= P(b \equiv a/1)P(1) + P(b \equiv a/2)P(2) + \dots + P(b \equiv a/n)P(n) \\ &= P(1) + 1/2P(2) + \dots + 1/nP(n) = \sum_{m=1}^n 1/mP(m). \end{aligned}$$

(d.h. wenn wir $m = 1, \dots, n$ Paare haben, die alle die fraglichen Merkmale aufweisen, ist die Wahrscheinlichkeit für das angeklagte Paar, auch das Täterpaar zu sein, gleich $1/m$).

Der Nenner läßt sich wie folgt umformen:

$$P(\geq 1) = 1 - P(0).$$

Die Wahrscheinlichkeit $P(0)$ ist das Produkt der einzelnen Wahrscheinlichkeiten, daß keines der n Paare die fraglichen Merkmale besitzt; $P(0)$ ist somit das Produkt $(1-p)$ mit sich selbst n Mal, d.h. $P(0) = (1-p)^n$.

Die Werte von $P(1), P(2), \dots, P(n)$ lassen sich durch die Binomialformel berechnen:

$$P(m) = \binom{n}{m} (p)^m (1-p)^{n-m}$$

$\binom{n}{m}$ heißen Binomialkoeffizienten, wobei gilt:

$$\binom{n}{m} = \frac{(n)_m}{m!} = \frac{(n)(n-1)\dots(n-m+1)}{1 \cdot 2 \cdot 3 \dots (m-1) \cdot (m)}$$

Wir erhalten somit:

$$\begin{aligned} P(b \equiv a / \geq 1) &= \frac{P(b \equiv a)}{P(\geq 1)} = \frac{1}{1-P(0)} \sum_{m=1}^n (1/m) P(m) \\ &= \frac{1}{1-(1-p)^n} \sum_{m=1}^n (1/m) \binom{n}{m} (p)^m (1-p)^{n-m} \end{aligned}$$

Wir können nun diese Formel auf den Collins-Fall anwenden. Wir ersetzen p durch die Wahrscheinlichkeit, zufällig ein Paar mit den genannten Merkmalen auszuwählen; diese ist nach Voraussetzung $1/12 \cdot 10^6$. n ersetzen wir durch die Anzahl der Paare, die zur Tatzeit in der Nähe des Tatorts gewesen sein könnten. Wenn n sehr groß und p sehr klein ist, können wir die Binomialverteilung durch die Poissonverteilung approximieren¹⁾.

Für $n \rightarrow \infty$ und sehr kleines p ist

$$\lim_{n \rightarrow \infty} P(m) = \lim_{n \rightarrow \infty} \binom{n}{m} (p)^m (1-p)^{n-m} = e^{-\lambda} \frac{\lambda^m}{m!} \quad \text{mit } \lambda = np.$$

Daraus ergibt sich die Formel:

$$P(b \equiv a / \geq 1) \approx \frac{1}{e\lambda - 1} \sum_{m=1}^{\infty} \frac{\lambda^m}{m \cdot m!} \quad 2)$$

Die Poisson-Approximation ergibt für $\lambda = 1$ den Wert 0.76696, für $\lambda = 1/10^2$ den Wert 0.99751³⁾

1) Vgl. die Schilderung der Kingston-Modelle auf S.93ff dieser Arbeit.

2) Die Ableitung dieser Formel ist bei Cullison, S.493, Fn.40, zu finden; vgl. auch Kingston, S.74.

3) Vgl. Kingston, S.74, Cullison, S.494.

Neben der Wahrscheinlichkeit für die Identität unter der Voraussetzung, daß es unter n Paaren mindestens ein Paar mit den fraglichen Merkmalen gibt, interessiert man sich gewöhnlich auch für die Wahrscheinlichkeit, daß es mehr als ein solches Paar unter der gleichen Voraussetzung gibt. D.h. man interessiert sich für die bedingte Wahrscheinlichkeit $P(>1/ \geq 1)$. Es ist klar: Je höher die Wahrscheinlichkeit ist, daß es mehr Paare mit den fraglichen Merkmalen gibt, desto kleiner ist die Wahrscheinlichkeit, daß das angeklagte Paar das Täterpaar ist.

Die Formel für die bedingte Wahrscheinlichkeit lautet:

$$P(>1/ \geq 1) = \frac{P(>1)}{P(\geq 1)}.$$

Den Nenner können wir wiederum umformen:

$$P(\geq 1) = 1 - P(0) = 1 - (1-p)^n.$$

Für den Zähler erhalten wir:

$$P(>1) = P(\geq 1) - P(1) = 1 - (1-p)^n - P(1) = (1 - (1-p)^n) - (np(1-p)^{n-1}).$$

Somit ist:

$$P(>1/ \geq 1) = \frac{(1 - (1-p)^n) - (np(1-p)^{n-1})}{1 - (1-p)^n} \quad 1)$$

Dies ist übrigens die Formel, anhand derer der Supreme Court den Schuldspruch im Collins-Fall revidiert hat. Mit $p = 1/12000000$ und $n = 12000000$ ($\approx$ die Bevölkerungszahl von Los Angeles) leitete das Gericht für $P(>1/ \geq 1)$ den Wert 0.42 ab; vgl. das Schlußwort²⁾

"Hence, even if we should accept the prosecution's figures without question, we would derive a probability of over 40 percent that the couple observed by the witnesses could be "duplicated" by at least one other equally distinctive interracial couple in the area, including a Negro with a beard and mustache, driving a partly yellow car in the company of a blonde with a ponytail".

1) Auch diese Formel können wir anhand der Poisson-Verteilung approximieren; es gilt:

$$P(>1/ \geq 1) \approx 1 - \frac{\lambda}{e\lambda - 1}; \text{ siehe Cullison, S.497, Kingston, S.76.}$$

2) 66 Cal. Rptr. 507, 438 P.2d 43, Cullison, S.497.

Vgl. dazu auch das Zitat aus Stripinis, S.65: "This glaring difference has often been used to put the use of probability theory into disrepute. However, it also points out the very great advantage of the use of mathematics. Without the use of mathematics, the prosecutor probably would have made a speech to the jury along the following line: "Don't you think it is an incredible coincidence that the accused couple should match the description of the guilty couple on all these points - yellow car, white roof, interracial couple, blonde hair, the ponytail, the moustache, and the beard?" The psychological effect on the jury would be about the same as the statement using probability theory..., but it is very unlikely that anyone would have challenged the factual content of the statement, saying that this was not such an incredible coincidence".

Die beiden abgeleiteten Formeln sind nicht unanfechtbar. Wir fassen hier kurz die Kritik Cullisons zusammen¹⁾.

1. Es wird die Möglichkeit zugelassen, daß die meisten und sogar alle Paare in der Population, die n Paare enthält, die fraglichen Merkmale besitzen. Dies ist jedoch eine unrealistische Annahme, die unserer Erfahrung widerspricht.
2. Das Vorgehen, daß der Ableitung der Formel für die Identität zugrunde liegt, entspricht einem Zufallsexperiment, das darin besteht, daß man ein Paar zufällig aus der Population, die aus n Paaren besteht, herausgreift. Dies entspricht jedoch nicht dem gewöhnlichen Vorgehen der Polizei, das nicht zufällig, sondern systematisch ist. Die Wahrscheinlichkeit für die Identität - die man eigentlich erhalten will - kann man aber nicht durch die Wahrscheinlichkeit für die Duplizierung (die die obige Voraussetzung der Zufälligkeit vermeidet) ersetzen, denn diese steht nur in einem sehr indirekten Zusammenhang mit der ersteren.

Was den Collins-Fall angeht, so ist auch Cullison der Auffassung, daß die Ableitung der Wahrscheinlichkeit $p = 1/12000000$ für das gemeinsame Vorkommen der fraglichen Merkmale fehlerhaft ist. Hier seine Kritik (S.504):

"To summarize: the prosecutor failed even to show that the probability p of a random couple having the six characteristics is $1/12,000,000$. This, for four reasons - (1) he merely guessed at the six individual probabilities he used in his calculations; (2) since some of the six characteristics may have been mutually interdependent, it was erroneous to use the simpler version of the multiplication rule for calculating the composite probability p ; (3) the six characteristics identify a narrower class of candidates for guilt than the eyewitness testimony would seem to justify; and (4) the prosecutor's argument does not take account of errors in his evidence. Of course this also means that the value $1/12,000,000$ for p cannot be used safely in our earlier formulae for calculating and estimating the probability of guilt - which means that the prosecutor was all the more off base, for these additional reasons, in claiming that the probability of guilt was $11.999.999/12,000,000$ ".

Aus dem obigen Zitat ersehen wir, daß neben der üblichen Kritik - daß die statistischen Daten nicht empirisch gestützt waren, nicht alle Befunde voneinander unabhängig waren, daß nicht alle Merkmale berücksichtigt wurden - wir bei Cullison einen neuen Einwand finden gegen die Art und Weise, wie der Anklagevertreter im Collins-Fall vorgegangen ist: So kritisiert er zu Recht, daß die Zuverläss-

1) Cullison, S.500ff.

sigkeit der Daten bzw. der Zeugenaussagen überhaupt nicht berücksichtigt wurde¹⁾²⁾

Hier wollen wir noch auf zwei Verbesserungsvorschläge der von Kingston und Culison verwendeten Methoden ganz kurz eingehen. Diese Vorschläge stammen einerseits von Smith und Charrow, andererseits von Stripinis³⁾

Smith und Charrow ersetzen die von Kingston vorgeschlagene Binomialverteilung durch die hypergeometrische Verteilung. Diese Methode besteht darin, daß man für eine Stichprobe aus einer bestimmten Population die Wahrscheinlichkeit für das Auftreten aller fraglichen Merkmale bei einer gewissen Anzahl von Individuen aus der Stichprobe bestimmt.

Da die Berechnung aufgrund der Formel für die hypergeometrische Verteilung mit Schwierigkeiten verbunden ist, schlagen Smith und Charrow eine Methode vor, die in der Bestimmung der oberen und der unteren Schranken für die Wahrscheinlichkeitsfunktion besteht. Sei λ der Erwartungswert für die Zahl der Individuen aus der Stichprobe, die alle fraglichen Merkmale besitzen, n sei die Zahl der Individuen in der zugrundeliegenden Population; m sei die Anzahl der in Frage kommenden Merkmale (diese werden als statistisch unabhängig angenommen) und $f_1, f_2, \dots, f_m$ die relativen Häufigkeiten ihres Auftretens in der Grundpopulation. Dann ist $\lambda = n \prod_{j=1}^m f_j$ (die Ableitung dieser Formel ist bei Smith&Charrow, S.558 zu finden).

Sei P_g die Abkürzung für die Wahrscheinlichkeit, daß der Verdächtige die fraglichen Merkmale (Spuren) am Tatort hinterlassen hat. Dann gilt für P_g :

$$\max \{0, 1 - \lambda\} \leq P_g \leq (1/(1+\lambda)) \quad (\text{ibid, S.558}).$$

Z.B. für $n = 10$, $m = 1$ (d.h. die Anzahl der fraglichen Merkmale ist 1) und $f_1 = 1/10$, folgt $\lambda = 1$ und es gilt $0 \leq P_g \leq 0.5$.

Als Mittelwert ergibt sich nach der von Smith&Charrow vorgeschlagenen Formel

$$(P_g = \frac{1-1/2\lambda^2}{1+\lambda} \quad \text{für } 0 \leq \lambda \leq 1 \quad \text{bzw. } P_g = \frac{1}{2(1+\lambda)} \quad \text{für } \lambda > 1):$$

$$P_g = 0.25 \quad (\text{der maximale Fehler beträgt auch } 0.25; \text{ vgl. S.559}).$$

1) Vgl. auch die Kritik von Kingston in "Probability and Legal Proceedings" bzw. S.114 dieser Arbeit.

2) Zur weiteren Kritik am Collins-Fall siehe W.B.Stoebuck, S.859ff. Zur Kritik des juristischen Mißbrauchs der Wahrscheinlichkeitstheorie siehe T.H.Liddle; allerdings unterläuft ihm der oben besprochene Fehler (d.h. er identifiziert die Wahrscheinlichkeit für das gemeinsame Vorkommen von Merkmalen mit der Wahrscheinlichkeit für die Identität); Liddle, S.277.

Dagegen richtig P.Sigal; aber auch er betrachtet (wie der Supreme Court) die Wahrscheinlichkeit für die Duplizierung, nicht jedoch die für die Identität, Sigal, S.183ff.

(Literaturangaben zu diesen Autoren sind auf S.112 Fn.1 dieser Arbeit zu finden).

3) R.L.Smith&R.P.Charrow, Upper and Lower Bounds for Probability of Guilt Based on Circumstantial Evidence, in: 70 J. of the American Stat. Ass. (1975), S.555-560. D.Stripinis, Probability Theory and Circumstantial Evidence: Implications from a Mathematical Analysis, in: 22 Jurimetrics J. (1981), S.59-82.

Anzumerken ist, daß die Wahrscheinlichkeit, die mit der Binomialverteilung abgeleitet werden kann, auch innerhalb dieser beiden Schranken liegt (vgl. *ibid*, S.559)¹⁾

Stripinis kritisiert den Ansatz von Kingston und Cullison, weil dieser, wie er schreibt, von der nicht korrekten Annahme ausgeht, daß "ein Individuum die fraglichen Merkmale hat"; stattdessen müsse es heißen "der Täter hat die fraglichen Merkmale" (vgl. Stripinis, S.66). Er ersetzt daher die bedingte Wahrscheinlichkeit für die Identität und Duplizierung durch die absoluten Wahrscheinlichkeiten. Seine Formeln lauten (mit "P(Id)" als Abkürzung für die Wahrscheinlichkeit der Identität und "P(Dup)" als einer solchen der Duplizierung):

$$P(\text{Id}) = \sum_{k=0}^{n-1} P(k) \left[\frac{1}{k+1} \right] \quad \text{und} \quad P(\text{Dup}) = P(>0) = 1 - P(0) = 1 - q^{(n-1)} \quad (\text{vgl. } \textit{ibid}, \text{ S.66}).$$

(n: die Zahl der Individuen in der in Frage kommenden Grundpopulation, die den Täter enthält;

n-1: eine solche ohne den Täter;

k: die Zahl der Individuen mit den gleichen Merkmalen wie der Täter in der Population, die den Täter nicht enthält;

q = 1-p: die Wahrscheinlichkeit, daß die fraglichen Merkmale bei einem zufällig gewählten Individuum nicht auftreten;

P(k): die Wahrscheinlichkeit, daß es k Individuen mit den fraglichen Merkmalen in der Grundpopulation ohne den Täter gibt).

1) Mit der Approximation der Binomialformel für große n ($p_g = P(X=0) = \left[1 - \frac{\lambda}{n}\right]^n$) erhalten wir für n = 10:

$$p_g = \left(1 - \frac{1}{10}\right)^{10} = \left(\frac{9}{10}\right)^{10} = 0.35.$$

Zur Kritik an Smith&Charrow und der Verteidigung von Kingston siehe D.J.Finney, Probabilities Based on Circumstantial Evidence, in: 72 Journal of the Am.Stat.Ass. (1977), S.316-18.

1.8. Finkelstein und Fairley¹⁾

Finkelstein und Fairley vertreten die Auffassung, daß die statistische Beurteilung allein nur in sehr seltenen Fällen überzeugend zur Identifizierung des Täters führt. Statt der herkömmlichen, nur statistischen Verfahren schlagen sie eine Methode vor, aufgrund derer man sowohl das statistische wie auch nichtstatistische Datum berücksichtigen kann: Wenn nichtstatistische belastende Befunde vorliegen, so kann das Bayes-Theorem²⁾ dazu benutzt werden, den Grad zu bestimmen, um welchen das zusätzliche statistische Datum den Verdacht vergrößert³⁾.

Auch Finkelstein und Fairley lehnen die im Collins-Fall benutzte Ableitung der Wahrscheinlichkeit für die Duplizierung als unkorrekt ab. Der Fehler dieses Ansatzes besteht ihrer Meinung nach darin, die Wahrscheinlichkeit p ($=1/12000000$) für die zufällige Wahl eines Paares mit den fraglichen Merkmalen aus einer Population ($\hat{=}$ Los Angeles) mit dem Parameter der Binomialverteilung gleichzusetzen, welcher wiederum auf die gleiche Population angewandt wird. Dies führt zum unsinnigen Resultat, daß auch dann, wenn die Population nur ein einziges Individuum mit den fraglichen Merkmalen enthält, man eine ziemlich hohe Wahrscheinlichkeit erhält (im Prozeß 0.42), daß es in dieser Population mehr als ein solches Individuum gibt.

Nach Finkelstein/Fairleys Meinung ist die Formel für die Duplizierung nur dann zulässig, wenn über die aktuelle Bevölkerungsstruktur von Los Angeles nichts bekannt ist, so daß - wenn $1/12000000$ die Wahrscheinlichkeit repräsentiert, daß ein zufällig ausgewähltes Paar die fraglichen Merkmale besitzt - etwa 40% aller möglichen Populationszusammensetzungen von Los Angeles wenigstens ein weiteres Paar mit den gleichen Merkmalen enthalten würden.

Finkelstein und Fairley kritisieren den bei der Lösung des Collins-Falls gewählten Ansatz auch aus einem anderen Grund: nämlich, weil er nur das statistische Datum berücksichtigt. Ihre Kritik an statistischen Methoden veranschaulichen sie durch das folgende Beispiel:

1) M.O.Finkelstein und W.B.Fairley, A Bayesian Approach to Identification Evidence, in: 83 Harvard Law Rev. (1970), S.489-517.

2) Zum Bayes-Theorem vgl. S.18 dieser Arbeit.
T.Bayes, An Essay Towards Solving a Problem in the Doctrin of Chance, Phil Tran. of the Royal Society (1763); Bayes schlug als erster vor, intuitive Wahrscheinlichkeitsurteile einerseits und Wahrscheinlichkeitsurteile, denen relative Häufigkeiten zugrunde liegen, andererseits mit Hilfe einer Methode zu kombinieren, die heutzutage als das Bayes-Theorem bekannt ist.

3) Von der Anwendung des Bayes-Theorems auf die juristischen Sachverhaltshypothesen spricht bereits Kaplan, Decision Theory and the Factfinding Process (1968), S.1083ff, allerdings ohne den Bayes-Ansatz dabei explizit zu formulieren. Vgl. auch Kingston, der das statistische Bayes-Modell (Kingstons viertes Modell; S.96 dieser Arbeit) in der Kriminalistik anwendet. Vgl. aber auch die Bayesianistische Interpretation der Booleschen Formel (S.85f dieser Arbeit) bzw. Schreibers Formel (S.109ff dieser Arbeit).

Angenommen, man findet in einem Strassengraben die Leiche einer Frau. Es ist bekannt, daß diese Frau am Abend vorher einen heftigen Streit mit ihrem Freund hatte. Es ist auch bekannt, daß er sie bereits mehrmals geschlagen hat. Bei der Untersuchung des Tatorts findet man die Tatwaffe, ein Messer, mit einem Fingerabdruck der rechten Hand, der einem solchen des Angeklagten ähnlich ist. Es handelt sich allerdings um einen sehr undeutlichen Fingerabdruck, so daß der herangezogene Experte lediglich sagen kann, daß nur jeder Tausendste einen solchen Fingerabdruck haben könne.

Wie Finkelstein und Fairley schreiben, hätte der Fingerabdruck nur einen geringen Beweiswert, wenn wir einen Ansatz wie im Collins-Fall benutzten. Denn wenn wir eine Population aus 100000 Individuen wählten, würde das heißen, daß 100 davon den gleichen Fingerabdruck hätten. Damit wäre diese Information wertlos. Dies widerspricht jedoch unserer Intuition, wonach gerade der gefundene Fingerabdruck den Angeklagten sehr stark belasten würde. Der mathematische Ansatz im Collins-Fall unterschätze offenbar diesen intuitiven Aspekt der Beweismittelbeurteilung.

Den Grund dafür sehen Finkelstein und Fairley darin, daß der frühere Ansatz alle Verdächtigen gleichermaßen behandelt. Das würde im vorliegenden Fall bedeuten, daß der Angeklagte nur zufällig als einer aus der Gruppe der 100 verdächtigen Personen mit dem gleichen Fingerabdruck gewählt wird unter Vernachlässigung aller anderen Informationen, die gerade seine Wahl nahelegen. Dies wäre jedoch nicht korrekt und entspräche nicht der Wirklichkeit; also müßten wir einen Weg finden, die angeklagte Person nicht als zufällig ausgewählt zu betrachten. Mit Finkelstein&Fairleys Worten (S.498):

"In our hypothetical case, the analysis of the palm print evidence should begin with the fact that defendant was far more likely to be guilty than someone selected at random. Consistent with this approach, it has been said that statistical evidence of the kind we have been considering should normally not be sufficient to support an identification unless accompanied by other evidence that would form the basis for a "prior" estimate of identity. This is an intuitive idea, but one that can be justified. We use Bayes' theorem for this purpose".

Finkelstein und Fairley sind also der Ansicht, daß das statistische Datum allein nicht ausreicht, daß auch das nichtstatistische Datum berücksichtigt werden muß und daß das Bayes-Theorem dafür die geeigneteste Methode ist, wobei mit Hilfe dieses Theorems vom nichtstatistischen informellen Datum (d.h. der Apriori-Einschätzung) unter Berücksichtigung des statistischen Datums (als Prämisse oder Bedingung) zur Aposteriori-Wahrscheinlichkeit übergegangen wird. Dies veranschaulichen Finkelstein und Fairley wie folgt (S.498ff):

Sei $P(G/F)$ die bedingte Wahrscheinlichkeit, daß der Angeklagte das Messer benutzt hat (d.h. daß es sich um seinen Fingerabdruck am Messer handelt) unter der Voraussetzung, daß ein seinem Fingerabdruck der rechten Hand ähnlicher am Messer festgestellt wird. $P(F/G)$ ist die Wahrscheinlichkeit, daß ein ähnlicher Fingerabdruck am Messer festgestellt wird, vorausgesetzt, daß es sich um seinen Fingerabdruck handelt.

G ist also das Ereignis, daß der Angeklagte das Messer benutzt hat (d.h. daß es sich um seinen Fingerabdruck am Messer handelt) und F das Ereignis, daß ein Fingerabdruck am Messer gefunden wird, der dem des Angeklagten (von dessen rechten Hand) ähnlich ist.

$P(F/\bar{G})$ ist die Wahrscheinlichkeit, daß ein ähnlicher Fingerabdruck am Messer festgestellt wird, vorausgesetzt, daß jemand anderer als der Angeklagte das Messer benutzt hat (d.h. daß es sich bei dem Fingerabdruck nicht um den des Angeklagten handelt).

Finkelstein und Fairley nehmen folgende Werte an:

$P(F/G) = 1$; $P(F/\bar{G}) = 1/1000$ (= die statistische Wahrscheinlichkeit für das Auffinden eines Individuums mit einem solchen Fingerabdruck in einer Population).

Die Bayes-Formel lautet:

$$P(G/F) = \frac{P(G)P(F/G)}{P(G)P(F/G) + P(\bar{G})P(F/\bar{G})}$$

Für $P(G)$ soll man nach Finkelstein&Fairley eine informelle Einschätzung aufgrund des vorhandenen nichtstatistischen Datums einsetzen.

Sei z.B. $P(G) = 0.01$ und $P(F/\bar{G}) = 0.001$, so erhalten wir $P(G/F) = 0.909$.

Kritik:

1. Wie wir sehen, bieten Finkelstein und Fairley eine sehr einfache Methode (das Bayes-Theorem) an, neben der intuitiven (aprioristischen) Einschätzung auch das statistische Datum zu berücksichtigen. Diese Methode ist übrigens auch dann anwendbar, wenn wir anstelle des statistischen Datums ein beliebiges Datum benutzen (wie es in der subjektivistischen Wahrscheinlichkeitstheorie der Fall ist). Überhaupt können wir mit dem Bayes-Theorem nicht nur, wie es die beiden Autoren tun, Identifizierungsprobleme lösen, sondern auch die Wahrscheinlichkeit für die Täterschaft bestimmen. Es ist daher erstaunlich, daß Finkelstein&Fairley es ablehnen, die Bayes-Methode auf andere als Identifizierungsprobleme anzuwenden¹⁾ Der Verdacht liegt nahe, daß sie fürchten, das für sie einigermaßen "sichere" Feld der Statistik zu verlassen und sich in das Gebiet

1) Vgl. Finkelstein&Fairley, The Continuing Debate over Mathematics in the Law of Evidence, in: 84 Harvard Law rev. (1971); S.1801, 1805.

der sog. "subjektiven Einschätzungen" zu begeben¹⁾

2. Im obigen Fall wird die Wahrscheinlichkeit für das Benutzen des Messers seitens des Angeklagten bestimmt. Es ist klar, daß auch dann, wenn der Angeklagte das Messer benutzt hat (d.h. sein Fingerabdruck darauf festgestellt ist), er nicht der Täter sein muß. Um das festzustellen (sei es mit dem Bayes-Theorem, sei es intuitiv), brauchen wir wiederum relevante Informationen.

Seien H die Abkürzung für die Hypothese, daß der Angeklagte der Täter ist, G dafür, daß er das Messer benutzt hat und F dafür, daß ein ähnlicher Fingerabdruck am Messer festgestellt worden ist. Die Wahrscheinlichkeit für die Hypothese, daß der Angeklagte der Täter ist, bedingt durch die Information, daß ein ähnlicher Fingerabdruck am Messer festgestellt worden ist, können wir auf eine Weise berechnen, die auch die unsichere Information berücksichtigt, daß der Angeklagte das Messer benutzt hat. Die Formel lautet:

$$P(H/F) = P(H/G \wedge F)P(G/F) + P(H/\bar{G} \wedge F)P(\bar{G}/F)^2$$

Nun verbrauchen Finkelstein & Fairley alle für die Feststellung der Schuld relevanten Informationen bereits für die Einschätzung der Apriori-Wahrscheinlichkeit bei dem Nachweis, daß der Angeklagte das Messer benutzt hat (also bei der Feststellung von $P(G/F)$, d.h. bei einem Beweismittel, das zur Beurteilung der Sachverhaltshypothese herangezogen werden soll, daß der Angeklagte der Täter ist). Will man also unter solchen Umständen die Hypothese, daß der Angeklagte der Täter ist, im obigen Beispiel probabilistisch beurteilen, so muß man entweder die gleichen Informationen doppelt zählen

1) Vgl. auch ihren - bei den Subjektivisten sicher auf heftige Ablehnung stoßenden - Versuch, den subjektivistischen Ansatz bei der Apriori-Beurteilung der Hypothese im vorangegangenen Beispiel frequentistisch "umzudeuten", S.504: "Although subjective probabilities can be used on this basis, we suggest that in the legal context they are likely to be interpreted as expressing a frequency, just as 'the chances of heads is one-half' expresses a frequency"; S.505: "Subjective probabilities reflect occurrences (as guilt) among events (cases) deemed similar because they generate similar degrees of belief". Vgl. allgemein auch: "There is in fact no significant difference between the subjective definition of probability applied to unique events and the classical definition based on relative frequency" (aus: Finkelstein, *Quantitative Methods in Law* (1978), S.64).

2) Vgl. die Bayes-Formel (eine Erweiterung der Bayes-Formel), S.34 dieser Arbeit.

(d.h. zum einen bei der Apriori-Beurteilung der Hypothese, daß der Angeklagte das Messer benutzt hat und zum anderen bei der einer solchen, daß der Angeklagte der Täter ist) oder, wenn man das vermeiden will, die Täterschaft des Angeklagten lediglich im Lichte der Information, daß er das Messer benutzt hat, intuitiv bestimmen (denn falls alle relevanten Informationen (für H) bereits bei der Bestimmung von $P(G/F)$ aufgebraucht worden sind, bleibt für die Berechnung von $P(H/G \wedge F)$ bzw. $P(H/\bar{G} \wedge F)$ nichts übrig). Beides ist aber entweder nicht korrekt oder nicht befriedigend.

Es ist außerdem wichtig, zwischen den Daten, die für die Beurteilung der Identität des festgestellten Fingerabdrucks relevant sind, und denen, die das nicht sind, oder allgemeiner: den Informationen, die für die Apriori-Beurteilung der Sachverhaltshypothese und denen, die für die Apriori-Beurteilung der Beweismittel-Hypothese herangezogen werden, zu unterscheiden. So ist im von Finkelstein & Fairley gewählten Beispiel die Tatsache, daß der Angeklagte mit dem Opfer Streit gehabt und es oft geschlagen hat, zwar für den Nachweis seiner Täterschaft, keineswegs aber für den der Identität der Fingerabdrücke relevant¹⁾ (Eine Information, die dafür vielleicht wichtig sein könnte, ist z.B. die Aussage von jemanden, ein ähnliches Messer beim Angeklagten gesehen oder ihm verkauft zu haben usw.). Da Finkelstein und Fairley uns nur die Daten für die Apriori-Beurteilung der Sachverhaltshypothese "Der Angeklagte hat die Tat begangen"²⁾, nicht dagegen für eine solche der von ihnen intendierten Identität des Fingerabdrucks präsentieren, ist das also ein Hinweis darauf, daß sie die Frage der Relevanz nicht entsprechend berücksichtigt haben³⁾

1) Denn das Hinterlassen von Fingerabdrücken an irgendwelchen Gegenständen hat mit Streitereien, Schlägereien und dergl., in denen diese Gegenstände nicht verwirklicht sind, nichts zu tun. Vertritt man im obigen Fall eine andere Auffassung, dann deshalb, weil man offenbar nicht mehr die Hypothese über die Identität des Fingerabdrucks beurteilt, sondern unbewußt bereits diejenige über die Täterschaft.

2) die sie gar nicht beurteilen wollen; vgl. S.1805, Fn.22 in: "The Continuing Debate...": "Professor Tribe discusses our hypothetical and finds it ambiguous as to whether the expert would use Bayes' theorem to direct the jurors to estimate the probability of defendant's guilt or the probability that the print came from the defendant... After observing that we use the phrase "prior probability of guilt", he concludes that the first interpretation was intended... Assuming that our hypothetical was ambiguous, what valid purpose is served by deliberately choosing the interpretation which makes the Bayesian approach unacceptable?... Nothing was said about using Bayesian techniques to prove defendant's guilt..." (Hervorhebung von mir).

3) Auf S.223ff dieser Arbeit werde ich einen ähnlichen Fall unter Vermeidung der oben angegebenen Schwierigkeiten aufgrund des subjektiven Charakters der Beweiswürdigung rekonstruieren.

1.9. Tribe¹⁾

Wie wir gesehen haben, gehört Finkelstein und Fairley der Verdienst, praktisch arbeitende Juristen mit dem Bayes-Theorem - einer Technik, mit deren Hilfe man in rationaler und konsistenter Weise mehrere Informationen zu einem Gesamtbild verarbeiten kann - bekannt gemacht zu haben. Anlaß für die Aufstellung dieser Formel war der, daß die bis dahin vorgeschlagenen Methoden ausschließlich statistischer Art und nur auf ein statistisches Datum anwendbar waren.

Das Bayes-Theorem erweist sich jedoch nicht nur, wie Finkelstein und Fairley meinen, für das Kombinieren des statistischen mit dem nichtstatistischen Datum von enormer Leistungsfähigkeit, sondern allgemein für die Informationsverarbeitung in bezug auf jedes quantifizierbare Datum (gleich ob statistischer oder nichtstatistischer Art)²⁾ Es ist daher keine Übertreibung zu behaupten, daß mit der Einführung des Bayes-Theorems in die Rechtspraxis die letztere erheblich an Wissenschaftlichkeit gewonnen hat. Der Vorschlag Finkelstein & Fairleys, das Bayes-Theorem für die Feststellung der Identität³⁾ nutzbar zu machen, stieß jedoch auf heftige Ablehnung seitens der Juristen, allen voran auf die von Laurence Tribe.

Wir wollen im folgenden Tribes Gedanken und Motivationen deshalb genauer betrachten, weil sie als programmatisch für die ablehnende Haltung des größten Teils der Juristen gegenüber wahrscheinlichkeitstheoretischen Methoden zum gegenwärtigen Zeitpunkt gelten können. Anschließend werden wir auf eine andere, nicht weniger polemische, Kritik an der Anwendung wahrscheinlichkeitstheoretischer Methoden in der Rechtspraxis eingehen. Es handelt sich dabei um den Aufsatz von Brilmayer und Kornhauser.

1) L.H. Tribe, Trial by Mathematics: Precision and Ritual in the Legal Process, in: 84 Harvard Law Rev. (1971), S.1329-1393; A Further Critique of Mathematical Proof, *ibid.*, S.1810-20.

2) Daß man auch das nichtstatistische Datum quantifizieren kann, bietet vielleicht für einen eingefleischten Objektivisten Schwierigkeiten, jedoch nicht für einen Subjektivist. Aber auch Finkelstein & Fairley, die eigentlich einen frequentistischen Standpunkt vertreten, können nicht umhin, die subjektive Einschätzung - nämlich die Quantifizierung der Anfangswahrscheinlichkeit der Hypothese im Lichte empirischer Befunde nichtstatistischer Art - zu verwenden. Auch wenn sie davon sprechen, daß subjektive Urteile in frequentistische Urteile übersetzbar seien (vgl. Finkelstein & Fairley, 1970, S.504f; auch S.126, Fn.1 dieser Arbeit), was angesichts der Komplexität und Vagheit solcher Urteile überhaupt nicht stimmen kann, bleibt die Tatsache bestehen, daß es sich dabei um subjektive Urteile aufgrund der persönlichen Überzeugung des Beurteilenden handelt.

3) Wie wir gesehen haben (S.125 dieser Arbeit), treten Finkelstein und Fairley mit einem viel geringeren Anspruch auf, als ihn das Bayes-Theorem ermöglicht: So wollen sie damit nur Probleme beurteilen, in denen es darum geht, festzustellen, ob gewisse empirische Befunde wie Blutspuren, Haare, Fingerabdrücke und dergl. vom Angeklagten stammen oder nicht. Sie wehren sich entschieden dagegen, das Bayes-Theorem auf die Beurteilung der Täterschaft bzw. auf sonstige induktive Argumentationen anzuwenden (vgl. S.127 Fn.2 dieser Arbeit).

Tribe vertritt die Auffassung, daß die Wahrscheinlichkeitstheorie "gefährlich", "dehumanisierend", "äußerst mißbräuchlich" und daher für die Rechtspraxis von geringem Nutzen ist¹⁾ sie führe nicht nur zu unakkuraten Ergebnissen, sondern zerstöre auch die traditionellen juristischen Werte²⁾

Tribe bestreitet zwar nicht, daß, wie er schreibt, das "gesamte"(!) Tatsachenwissen "statistisch" und jeder juristische Beweis "probabilistisch" ist, wendet sich jedoch gegen eine explizite Wahrscheinlichkeitstheoretische Beurteilung juristischer Sachverhalte³⁾

Auch warnt er, daß die Wahrscheinlichkeitstheorie wegen ihres Anspruchs auf größere Präzision und Zuverlässigkeit für Juristen viel zu attraktiv sein könnte, so daß diese - da sie i.d.R. wenig von der Wahrscheinlichkeitsrechnung verstünden - auch unrichtige Resultate kritiklos zu akzeptieren bereit wären. Als Illustration führt er die Dreyfuß-Affaire an, wo im Prozeß angewandte, zweifelhafte Beweismethoden zu ungerechtfertigten Schlußfolgerungen Anlaß gaben, die die Richter dermaßen beeindruckten, daß Dreyfuß aufgrund dieser Beweisführung verurteilt wurde⁴⁾⁵⁾

1) Vgl. Tribe, S.1330f.

2) Vgl. *ibid.*, S.1330, 1376.

Dazu ist nur folgendes anzumerken: Da die Wahrscheinlichkeitstheorie wie jede mathematische Disziplin wertneutral ist und lediglich ein Werkzeug darstellt, um in konsistenter Weise von beliebigen Behauptungen zu anderen Behauptungen überzugehen (wobei solche Behauptungen natürlich auch inhaltlich falsch sein können, weil die Aufstellung solcher Behauptungen letztlich von Menschen abhängt), sollte man daher gerechterweise bei der Ablehnung eigentlich bei den Menschen anfangen, die dazu neigen, solche Methoden zu mißbrauchen. In dieser Hinsicht ist die Situation mit der formalen Logik vergleichbar; dort kann man logisch korrekte Schlüsse aus inhaltlich sinnlosen Aussagen ziehen, wobei das Ergebnis auch inhaltlich sinnlose Aussagen sind. Es fällt jedoch heutzutage kaum mehr jemandem ein, die Logik deswegen abzulehnen.

3) Tribe, S.1330, Fn.2.

4) Vgl. *ibid.*, S.1332ff. Auf die Fehler in der dortigen Beurteilung hat bereits der Mathematiker Henri Poincaré, der den Dreyfuß-Fall erstmals analysiert hat, hingewiesen. Vgl. Mode, S.628-640; S.639 und Cullison, S.510, die diesen Fall als Illustration für die Unfähigkeit des Gerichts anführen, mit statistischen Daten umzugehen. Die von Tribe geäußerte Warnung scheint angesichts der weitverbreiteten Abneigung der Juristen gegenüber mathematischen Methoden eigtl. eher überflüssig zu sein.

5) Auch andere von Tribe angeführten Kritiken (Collins-Fall, McCormick usw.), vgl. Tribe, S.1334ff, wurden bereits vor Tribe von Anhängern der Wahrscheinlichkeitstheorie wie Kingston, Cullison vorgebracht (die im Unterschied zu Tribe diese Kritiken nicht gegen die Wahrscheinlichkeitstheorie, sondern gegen den Dilettantismus der meisten solche Methoden anwendenden Juristen gerichtet haben).

Weiter führt Tribe die sog. traditionellen Einwände gegen die Wahrscheinlichkeitstheorie an.

Typisch für Tribe ist dabei, daß er seine Einwände an Beispielen demonstriert, bei deren Beurteilung lediglich das statistische Datum benutzt wird¹⁾ wie etwa an folgendem²⁾

Der Kläger wurde fahrlässig von einem blauen Bus angefahren. Die Frage ist, ob der Bus dem Angeklagten gehört oder nicht. Der Kläger kann beweisen, daß dem Angeklagten 80% aller blauen Busse in der Stadt gehören. Also folgt nach der Wahrscheinlichkeitstheorie (in Tribes Interpretation), daß der Angeklagte mit der Wahrscheinlichkeit von 80% am Unfall schuldig ist. Hier folgt die typische Frage Tribes: Welche Wirkung (Sinn), wenn überhaupt, soll ein solcher Beweis haben? (Aus subjektivistischer Sicht würde man einem solchen Beweis nicht mehr oder weniger Bedeutung beimessen, als es intuitiv gesehen gerade der Fall ist).

Diese Einwände sind³⁾

1. Man nimmt an, daß der statistische Wahrscheinlichkeitsbegriff nur auf die zukünftigen Fälle anwendbar ist und daher für die Beurteilung juristischer - die ja historischer Art sind - nicht geeignet ist.
2. Der statistische Wahrscheinlichkeitsbegriff erlaubt nur allgemeine Aussagen, jedoch nicht solche über Einzelereignisse.
3. Die mathematischen Daten allein können nicht einen juristischen Fall entscheiden.

Während der erste Einwand nicht richtig ist⁴⁾ ist der zweite - nicht nur aus juristischer, sondern auch aus subjektivistischer Sicht - korrekt. Tribe vertritt jedoch eine andere Ansicht, wenn er schreibt⁵⁾ daß es möglich ist, vom Wahrscheinlichkeitsbegriff im Sinn der relativen Häufigkeit zum Wahrscheinlichkeitsbegriff im Sinn des subjektiven Glaubens überzugehen, indem man - hier nimmt Tribe bezug auf Savage - auf das fragliche Ereignis Wetten abschließt, im obigen Beispiel etwa, wenn der Richter bereit wäre, einen der Wahrscheinlichkeit 0.8 entsprechenden Einsatz zu leisten, daß der Kläger von einem blauen Bus, der dem Angeklagten gehört, angefahren wurde. Tribe mißversteht anscheinend den subjektivistischen Ansatz, wenn er meint, das Problem der Einzelfallwahrscheinlichkeit sei gelöst, wenn man zwar den statistischen Wahrscheinlichkeitsbegriff beibehält, diesen aber zur Not mit dem Glaubensgrad gleichsetzt. Denn entweder handelt es sich um den Begriff

1) Tribe, S.1339ff.

2) *ibid.*, S.1340f.

3) *ibid.*, S.1344ff.

4) Auch Tribe ist dieser Auffassung, vgl. Tribe, S.1345f.

5) *ibid.*, S.1347f.

der statistischen Wahrscheinlichkeit (dann kann man nicht mehr ohne weiteres über Einzelereignisse sprechen) oder man wählt von vornherein den subjektivistischen Ansatz (dann braucht man nicht, wie im obigen Beispiel, mit der Wahrscheinlichkeit 0.8 daran zu glauben, daß der Bus dem Angeklagten gehört)¹⁾

Was den dritten Einwand betrifft, daß mathematische Daten allein den juristischen Fall nicht entscheiden können²⁾, so läßt sich dazu folgendes sagen:

Interpretiert man, wie Tribe das tut, nur statistische Daten als mathematische, d.h. als Daten, die quantifizierbar sind³⁾, so hat

1) In einer anderen Hinsicht hat Tribe Recht: So ist es möglich, innerhalb der subjektivistischen Wahrscheinlichkeitstheorie auch über statistische Ereignisse zu sprechen (vgl. S. 46f dieser Arbeit). Nur muß man dann auch den subjektivistischen Ansatz bei der Beurteilung der Fälle wählen und nicht, wie Tribe es tut, zuerst die relative Häufigkeit feststellen und dann versuchen, sie subjektivistisch zu rechtfertigen (vgl. Tribes Repräsentation juristischer Fälle auf S. 1339ff in eindeutig statistischer Ausdrucksweise); vgl. *ibid.*, S. 1347f: "In the context of the bus case, if a person knew only that eighty percent of all blue buses are operated by the defendant (It is possible to know only this fact at the outset of the trial -...- Anm. in Fn. 59), and if he had to bet one way or the other, he should be as willing to bet that the bus involved in this case was defendant's as he would be willing to bet that a card chosen at random from B_{80} (B_{80} ist eine Urne, die 80 Kärtchen enthält, die mit "wahr", und 20, die mit "falsch" markiert sind, -Anm. von mir) would be marked "True". This is what it means to say that, for him, the subjective probability of the defendant's liability... equals .8".

Auch wenn man weiß, daß 80% aller blauen Busse in der Stadt dem Angeklagten gehören, und dies die einzige Information ist, braucht man nicht mit 80% Wahrscheinlichkeit daran zu glauben, daß der fragliche Bus dem Angeklagten gehört. Man kann mit jeder beliebigen Wahrscheinlichkeit daran glauben: etwa mit der Wahrscheinlichkeit 0, wenn man diese statistische Information für die Beurteilung dieses konkreten Falls als irrelevant ansieht, oder aber auch mit der Wahrscheinlichkeit 0.8, wenn man die statistische Information für relevant und für die Beurteilung des obigen konkreten Falls für völlig ausreichend hält). Darin (d.h. daß man die intuitive Haltung wahrscheinlichkeitstheoretisch nachvollziehen kann) und nicht in der Rechtfertigung der statistischen Wahrscheinlichkeit besteht eigtl. der Sinn des subjektivistischen Ansatzes).

2) Auch Tribe ist dieser Auffassung; vgl. Tribe, S. 1360.

3) Vgl. *ibid.*, S. 1350. Tribe identifiziert mathematische Daten schlichtweg mit statistischen, auch wenn er bisweilen einen Versuch unternimmt, die Statistik subjektivistisch zu rechtfertigen. Gelegentlich ist Tribe allerdings sogar der Ansicht, daß in manchen Fällen das statistische Datum allein ausreichen dürfte; vgl. Tribe, S. 1349, Fn. 65: "Of course, if plaintiff can satisfactorily account for the evidentiary omission, the statistical evidence alone might well suffice".

dieser Einwand Berechtigung. Ist man, wie das aufgrund des subjektivistischen Ansatzes der Fall ist, der Ansicht, daß auch nicht-statistische Daten quantifizierbar und damit alle induktiven Argumente formal rekonstruierbar sind, so wird man diesen Einwand nicht akzeptieren¹⁾

Im folgenden wollen wir uns mit Tribes Kritik am Bayesianistischen Ansatz von Finkelstein&Fairley beschäftigen.

Tribe schreibt (S.1354f):

"...although their analysis is both intriguing and illuminating, neither the technique proposed by Finkelstein and Fairley, nor any other like it, can serve the intended purpose at an acceptable cost".

Tribe wendet sich gegen die Lösung des von Finkelstein und Fairley besprochenen Falls mit Hilfe des Bayes-Theorems. Es ist interessant, seine Einwände näher zu betrachten²⁾ Diese sind:

1. Die Richter können ihre subjektiven Einschätzungen nicht quantifizieren, d.h. sie können dem, was Finkelstein&Fairley als Apriori-Wahrscheinlichkeit bezeichnen, überhaupt keine Zahl zuordnen, die korrekt ihre intuitive Einschätzung wiedergibt. Mit Tribes Worten (S.1359):

"Because the Finkelstein-Fairley technique thus compels the jury to begin with a number of the most dubious value, the use of that technique at trial would be very likely to yield wholly inaccurate, and misleading precise, conclusions".

1) Einen weiteren Einwand Tribes wollen wir hier nur kurz erwähnen (vgl. Tribe, S.1350), nämlich den, daß die Wahrscheinlichkeitstheorie den Beweis "mit an Sicherheit grenzender Wahrscheinlichkeit" (wobei dies im Sinne, daß "nur erkenntnistheoretische Zweifel bestehen", vgl. S. 58, Fn.1 dieser Arbeit) zu verstehen ist - Anm. von mir) nicht garantiere. Dies betrifft jedoch nicht nur die statistische Wahrscheinlichkeitstheorie, sondern jede Form der Präzisierung von intuitiven Überlegungen.

2) Es sei hier eine Ungenauigkeit erwähnt, die ihrerseits Tribe zu vielen Einwänden Anlaß gab.

Tribe nimmt - woran Finkelstein&Fairley eigtl. durch manche ihre mißverständlichen Formulierungen teilweise auch Schuld tragen - fälschlicherweise an, daß Finkelstein und Fairley den Ausdruck "used the knife" im Sinn "used the knife to kill his girlfriend" (vgl. Tribe, S.1356; vgl. S.127, Fn.2 dieser Arbeit) verwenden. Der Nachweis mittels des Bayes-Theorems wäre dann, ganz entgegen der Absicht von Finkelstein&Fairley, der der Täterschaft des Angeklagten und nicht der der Annahme, daß der Fingerabdruck am Messer der des Angeklagten ist (Auf die Einwände Tribes, die auf diesem Mißverständnis beruhen, werde ich nicht eingehen).

Dieser Einwand scheint nicht so sehr gegen die Anwendung des Bayes-Theorems gerichtet zu sein, sondern eher gegen die Quantifizierung von subjektiven (qualitativen) Urteilen. Es ist jedoch fraglich, ob solche Quantifizierungen in der Tat (wie Tribe das im obigen Zitat behauptet) nicht akkurat sind bzw. weniger akkurat als das verschwommene, (nach Meinung Tribes) nicht objektivierbare und daher nicht nachprüfbare "Rechtsgefühl" des Richters¹⁾

Im übrigen ist es gar nicht so schwer, von einem qualitativen Urteil der Art "Ich schätze die Glaubwürdigkeit des Zeugen hoch" oder "Die Beweismittel sind meiner Auffassung nach für den Sachverhalt nicht ausreichend" usw. zu entsprechenden numerischen Werten überzugehen. Eine Möglichkeit (die Savage vorschlägt) erwähnt ja Tribe selber²⁾

2. Das statistische Datum "erdrückt" die Apriori-Einschätzung, so daß am Ende nicht das nichtstatistische, sondern das statistische Datum über den Fall entscheidet³⁾

Auch dieser Einwand richtet sich eigentlich nicht gegen das Bayes-Theorem selbst, sondern gegen seine Anwendung auf das statistische Datum. Denn es kann in der Tat passieren, daß das statistische Datum, wenn es genügend stark und das sonstige nichtstatistische Datum hinreichend schwach ist, bei der Urteilsbildung den Ausschlag gibt. Ob man solche Fälle (d.h. Fälle, wo das statistische Datum stärker ist als das nichtstatistische) ausschließen soll oder nicht, muß die Rechtsprechung entscheiden⁴⁾ (denn es handelt sich hier nicht mehr um eine wahrscheinlichkeitstheoretische, sondern vielmehr um eine methodologische Frage).

1) Da in dieser Arbeit auf den Nutzen und die Ziele der Quantifizierung vielerorts eingegangen wird (vgl. S. 61f, 199f und v.a. den zweiten Teil dieser Arbeit), sei hier nur ganz kurz erwähnt, daß diese erstens in der Präzisierung des subjektiven, intuitiven Urteils (bzw. Gefühls), zweitens in dessen 'Objektivierung' (zwecks der intersubjektiven Nachvollziehbarkeit der Urteilsbildung) und drittens in der Rationalisierung der juristischen Entscheidungsbildung bestehen.

Ich werde im nächsten Paragraphen Ergebnisse einiger empirischer Untersuchungen schildern, die die Behauptung widerlegen, daß Juristen ihre qualitativen, intuitiven Urteile nicht durch entsprechende numerische Werte einigermaßen genau wiedergeben können.

Zur Kritik am "Rechtsgefühl" des Richters vgl. S. 73 dieser Arbeit.

2) Tribe, S. 1347.

3) *ibid.*, S. 1359ff.

4) Typische Fälle dieser Art sind z.B. solche, in denen es um den Nachweis der Vaterschaft bzw. Nichtvaterschaft geht.

3. Die sog. "weichen" Variablen (d.h. Variablen, die nicht quantifizierbar sind) werden unterdrückt¹⁾

Als Beleg dafür führt Tribe die Tatsache an, daß Finkelstein und Fairley $P(F/G) = 1$ ansetzen (d.h. die Wahrscheinlichkeit, daß ein ähnlicher Fingerabdruck am Messer gefunden wird unter der Voraussetzung, daß der Angeklagte das Messer benutzt hat).

Interpretiert man den Ausdruck "Der Angeklagte hat das Messer benutzt" wie Tribe es tut, als "Der Angeklagte hat das Messer benutzt, um seine Freundin zu töten", so kann dieser Wert natürlich kleiner sein, etwa wenn die Möglichkeit besteht, daß der Angeklagte Handschuhe trug bzw. seine Fingerabdrücke abwischen konnte usw.²⁾ Um das jedoch im konkreten Fall beurteilen zu können, müssen wir mehr über den fraglichen Sachverhalt wissen³⁾ Erweist es sich, daß irgendwelche Tatsachen gegen den Wert 1 für $P(F/G)$ sprechen, so ist es kein Problem, diesen durch einen der Auffassung des Beurteilenden angemesseneren zu ersetzen.

Gegen diesen Einwand können wir außerdem die folgenden zwei Argumente anführen: Erstens, es ist nicht Aufgabe eines Theoretikers, der eine Berechnungs- (bzw. Beurteilungs-) Methode vorschlägt, auch die Werte zu rechtfertigen, die lediglich zum Zweck der Veranschaulichung der Funktionsweise dieser Methode dienen. Es ist vielmehr die Aufgabe des praktisch arbeitenden Juristen bei der Quantifizierung alle in Frage kommenden (bzw. alle in seinen Augen relevanten) Variablen zu berücksichtigen⁴⁾

Zweitens, es wird unerschwerlich die (bereits kritisierte) Auffassung von Tribe vertreten (etwa in der Behauptung, daß die sog. "weichen Variablen" nicht quan-

1) Vgl. Tribe, S.1361f.

2) Vgl. dazu S.132Fn.2 dieser Arbeit. Interpretiert man dagegen diesen Ausdruck im Sinne von Finkelstein&Fairley als "Es handelt sich bei dem Fingerabdruck am Messer um den des Angeklagten", so ist der obige Wahrscheinlichkeitswert plausibel, denn aus der Tatsache, daß es sich um den Fingerabdruck des Angeklagten am Messer handelt, folgt ja, daß dieser sich selbst ähnlich ist.

3) Tribe irrt sich mit seiner Meinung, solche Überlegungen könnten nicht berücksichtigt werden. Gerade die Zuordnung von $P(F/G)$ seitens des Gerichts wird Aufschluß darüber geben, welche Bedeutung der Richter dem Fingerabdruck im Lichte anderer Informationen und Kenntnisse beimißt.

4) Finkelstein&Fairleys Absicht war es, zu zeigen, wie der Prozeß der Informationsverarbeitung funktioniert. Die Rechtfertigung der numerischen Werte gehört dagegen in einen Kontext, für den nicht mehr der Theoretiker, sondern der praktisch arbeitende Jurist verantwortlich ist.

Wie wir sehen, bringt Tribe zwei Aspekte der rationalen Begründung durcheinander: zum einen die Wahl der Methoden und zum anderen die Rechtfertigung der Annahme bzgl. der Wahl von bestimmten Wahrscheinlichkeitswerten. Die Kritik an letzterem ist ihm Vorwand, ersteres anzulehnen.

tifizierbar sind), daß nur das statistische Datum numerisch ausdrückbar ist, d.h. es nur eine einzige Möglichkeit gibt, juristische Sachverhalte probabilistisch zu beurteilen, nämlich aufgrund der statistischen Wahrscheinlichkeitstheorie.

Daß der Bayes-Ansatz in den Augen Tribes "die weichen Variablen" unterdrücken soll, ist um so verwunderlicher, als Tribe auf S.1364 selbst zeigt, daß diese Methode funktioniert (auch dann), wenn man die Zahlen anders, seiner Meinung nach, adäquater wählt.

Da uns die subjektive Wahrscheinlichkeitstheorie die Möglichkeit gibt, auch mit dem nichtstatistischen Datum umzugehen, entbehrt die folgende Behauptung Tribes jeglicher Berechtigung (S.1365): "On the contrary, the excessive weight that will thereby be given to those factors that can most easily be treated mathematically indicates that, on balance, more mistakes may well be made with partial quantification than with no quantification at all".

4. Ein und dasselbe Datum wird doppelt gezählt: zum einen bei der Bestimmung der Apriori-Wahrscheinlichkeit der Sachverhaltshypothese, zum anderen bei der wahrscheinlichkeitstheoretischen Beurteilung dieser Hypothese aufgrund des Datums. Mit Tribes Worten (S.1366):

"Essential to the application of Bayes' Theorem to derive $P(X/E)$ from $P(X)$ is that the trier be able somehow to make a prior estimate of $P(X)$, the probability of the disputed proposition X . If that estimate is arrived at after knowing or even suspecting E , then to use the information provided by E to refine the estimate through Bayes' Theorem to obtain $P(X/E)$ would obviously involve counting the same thing twice".

Dies scheint wiederum nicht gegen den Bayesianistischen Ansatz gerichtet zu sein, sondern gegen den nicht korrekten Gebrauch solcher Methoden durch nichtqualifizierte Juristen¹⁾

5. Man ist gezwungen, quantitative Schätzungen vorzunehmen, auch wenn das Verfahren noch lange nicht abgeschlossen ist (vgl.S.1368):

At least in criminal cases, and perhaps also in civil cases resting on allegation of moral fault, further difficulties lurk in the very fact that the trier is forced by the Finkelstein-Fairley technique to arrive at an explicit quantitative estimate of the likely truth at or near the trial's start, or at least before some of the most significant evidence has been put before him".

1) Vgl. auch Tribe, S.1367: "The accurate application of Bayes' Theorem along the lines proposed by Finkelstein und Fairley necessarily assumes that the evidence E of quantifiable probative value can be made independent of the prior suspicion, but in most trials the two will be hopelessly enmeshed". Auch dies spricht nicht so sehr gegen Finkelstein & Fairley, die ja gezeigt haben, daß die Apriori-Beurteilung im Lichte bestimmter (spezifizierbarer), vor Gericht gebrachter Informationen geschehen soll, sondern gegen die Gerichtspraxis selbst, in der man, Tribes Meinung zufolge, nicht imstande ist, in bezug auf die Informationsverarbeitung zu differenzieren.

So, wie Tribe es beschreibt, muß es nicht sein. Man muß nicht die Schätzungen bzw. Berechnungen am Anfang des Prozesses vornehmen; man kann das (und es empfiehlt sich sogar) am Ende tun, wenn alle Daten bekannt sind. Aber auch bei provisorischen Berechnungen in irgendeinem früheren Stadium kann man fragen, was dabei so schlimm ist. Denn diese können jederzeit korrigiert werden, falls neue Informationen hinzukommen.

Auch die Apriori-Schätzung der Hypothese braucht nicht im Lichte dessen erfolgen, was zu Anfang des Prozesses präsentiert wird. Das können beliebige Beweismittel sein. Bei der wahrscheinlichkeitstheoretischen Feststellung des Gesamtgewichts aller relevanten Informationen spielt es keine Rolle, in welcher zeitlichen Abfolge die Beweismittel dem Gericht vorgetragen werden. So kann man der aprioretischen (oder intuitiven) Beurteilung auch ein Hauptbeweismittel zugrunde legen oder überhaupt das gesamte vorhandene Datum (was sich eigentlich vom herkömmlichen traditionellen Vorgehen nur dadurch unterscheidet, daß wir hier dieser Art Gesamturteil einen numerische Wert zuordnen)¹⁾ In diesem Fall bleibt, falls keine zusätzlichen Informationen hinzukommen, diese aprioretische (intuitive) Beurteilung auch die Endbeurteilung des Falls.

6. Man nimmt, bevor das Verfahren abgeschlossen ist, an, daß der Angeklagte schuldig ist, was der juristischen Konvention "man ist solange unschuldig bis man verurteilt wird" widerspricht²⁾

Darauf ist zu antworten, daß die Apriori-Einschätzung - insbesondere, wenn sie im Lichte nicht aller vor Gericht gebrachten Informationen geschieht - nicht mehr als ein provisorisches (und vermutlich nur dem Beurteilenden bekanntes) Urteil darstellt, das aufgrund anderer Informationen bestärkt oder modifiziert werden kann.

Außerdem ist es ein Unterschied, ob man die Schuld nur vermutet (was die Apriori-Einschätzung ausdrückt) oder sie als sicher bzw. "mit an Sicherheit grenzender Wahrscheinlichkeit" annimmt, bevor die Verurteilungsformel ausgesprochen wird. Daß eine Vermutung bzw. ein Verdacht gegen den Angeklagten besteht, lange bevor das

1) Vgl. dazu die experimentelle Methode von Simon & Mahan, die in der Quantifizierung des intuitiven Gesamturteils besteht. Hingegen kombinieren Marshall & Wise das Apriori-Urteil mit neuen Daten, um auf diese Weise die Aposteriori-Wahrscheinlichkeit der Hypothese zu erhalten. (Überdies und ähnliche empirische Untersuchungen berichten wir im nächsten Paragraphen).

2) Tribe, S.1370.

endgültige Urteil gesprochen ist, wird wohl niemand bestreiten. Denn dem Angeklagten würde kein Prozeß gemacht, wenn man von seiner Unschuld von vornherein wirklich überzeugt wäre.

Darüber hinaus ist es falsch anzunehmen, daß der Richter erst mit der Urteilsverkündung von der Schuld des Angeklagten überzeugt wird und nicht schon viel früher seine Vermutungen diesbezüglich anstellt, die zu Ende des Prozesses eine bestimmte endgültige Gestalt annehmen¹⁾

Und schließlich, was ändert sich schon an der Gültigkeit der oben zitierten juristischen Formel, wenn der Richter während des Prozesses irgendeine Wahrscheinlichkeit für die Schuld des Angeklagten annimmt. Denn vor der Öffentlichkeit – und darauf kommt es primär (auch Tribe) an – bleibt der Angeklagte unschuldig, solange die Verurteilung nicht ausgesprochen ist²⁾

Am Ende seiner Kritik am Bayesianistischen Ansatz erfahren wir übrigens noch, warum ein Gebrauch des Bayes-Theorems, oder allgemeiner, mathematischer Methoden, zur "Dehumanisierung" der Jurisprudenz führe. Der Grund dafür besteht nach Tribe darin, daß er "den Charakter des Gerichtsprozesses verändert"³⁾ eines Prozesses, dessen Aufgabe "nur zum Teil im Auffinden der historischen Wahrheit" besteht, denn jener ist, was nicht weniger wichtig ist, auch ein Ritual – "a complex pattern of gestures comprising what Henri Hart und John McNaughton once called "society's last line of defense in the indispensable effort to secure the peaceful settlement of

1) Zu diesem Punkt vgl. auch S.32 dieser Arbeit.

2) Wir werden von Tribe auf S.1371f darüber aufgeklärt, wozu in Wirklichkeit diese juristische Konvention gut ist: "Indeed, I have not proposed that anyone deceive either himself or another about the factual underpinning of the presumption of innocence, but only that worthwhile values served by that presumption as a normative standard might be harder to secure if the probability of guilt became a matter for precise and explicit assessment and articulation early in the typical criminal trial. The point, then, is not that any factual truth should be concealed or even obscured, but only that one need not say everything all at once in order to be truthful, and that saying some things in certain ways and at certain times in the trial process may interfere with other more important messages that the process should seek to convey and with attitudes that it should seek to preserve.

3) Tribe, S.1375f.

social conflicts"²⁾ Dem ist nicht mehr hier hinzuzufügen, als daß Tribes Fußnote. 148 auf S.1375 ebensowenig überzeugend ist:

"Suppose, for example, that each of three items of evidence, E_1, E_2 and E_3 , has the effect of increasing a prior 1 percent suspicion of guilt ($P(\bar{X})=.01$) ten-fold, so that $P(X/E_1) = P(X/E_2) = P(X/E_3) = .1$. If E_1, E_2 and E_3 are conditionally independent of X and not- X ,... then it turns out that $P(X/E_1 \& E_2 \& E_3)$ is in excess of .93, a result that might be counter-intuitive for many laymen".

" E_1, E_2 and E_3 are conditionally independent of X and not- X " heißt bei Tribe, daß gilt: $P(E_1 \& E_2 \& E_3/X) = P(E_1/X)P(E_2/X)P(E_3/X)$ und $P(E_1 \& E_2 \& E_3/\text{not-}X) = P(E_1/\text{not-}X)P(E_2/\text{not-}X)P(E_3/\text{not-}X)$, vgl. Fn.75, S.1352.

Es gilt somit:

$$\begin{aligned} P(X/E_1 \& E_2 \& E_3) &= \frac{P(E_1 \& E_2 \& E_3/X)P(X)}{P(E_1 \& E_2 \& E_3/X)P(X) + P(E_1 \& E_2 \& E_3/\bar{X})P(\bar{X})} \\ &= \frac{P(E_1/X)P(E_2/X)P(E_3/X)P(X)}{P(E_1/X)P(E_2/X)P(E_3/X)P(X) + P(E_1/\bar{X})P(E_2/\bar{X})P(E_3/\bar{X})P(\bar{X})} \end{aligned}$$

Nach Voraussetzung ist nur der Wert $P(X) = 0.01$ bekannt (die anderen Werte $P(X/E_i) = 0.1$, $i = 1, 2, 3$, sind für die Berechnung mit der Formel irrelevant).

Um den von Tribe behaupteten Wert von 0.93 abzuleiten, muß das folgende gelten (y steht für $P(E_i/\bar{X})$ und x für $P(E_i/X)$, $i = 1, 2, 3$):

$$0.93 = \frac{x^3 0.01}{x^3 0.01 + y^3 0.99} \quad \text{bzw.} \quad 0.93(0.01x^3 + 0.99y^3) = 0.01x^3.$$

Daraus folgt: $x^3 = 1315y^3$.

Das heißt: das Verhältnis des Produkts aus $P(E_1/X)$, $P(E_2/X)$, $P(E_3/X)$ zum Produkt aus $P(E_1/\bar{X})$, $P(E_2/\bar{X})$, $P(E_3/\bar{X})$ ist 1315 zu 1; ist etwa $P(E_1/X) = 1$, so erhalten wir $P(E_1/\bar{X}) \approx 0.09$.

Angesichts $P(E_1/X) = P(E_2/X) = P(E_3/X) = 1$ (d.h. alle Informationen sind, falls die Hypothese stimmt, völlig zuverlässig) und $P(E_1/\bar{X}) = P(E_2/\bar{X}) = P(E_3/\bar{X}) \approx 0.09$ (d.h. alle Informationen sind, falls die Hypothese nicht stimmt, nur zu 0.09 unzuverlässig) ist der Wahrscheinlichkeitswert 0.93 aufgrund dieser insgesamt sehr zuverlässigen Informationen überhaupt nicht "kontra-intuitiv".

(Der Übergang von $P(X) = 0.01$ zu $P(X/E_i) = 0.1$, $i = 1, 2, 3$, und von $P(X/E_1)$ zu $P(X/E_1 \& E_2 \& E_3) = 0.93$ läßt sich folgendermaßen interpretieren: Jedes Datum macht die Hypothese um zehnmal und alle Daten zusammen um über neunzigmal wahrscheinlicher bzw. um über neunmal wahrscheinlicher als jedes einzelne Datum. Es ist schwer zu sehen, was an einem solchen Übergang "kontra-intuitiv" ist)²⁾

1) Tribe, S.1376.

2) Zur weiteren Kritik an Tribe vgl. Stripinis, Probability Theory and Circumstantial Evidence: Implications from a Mathematic Analysis, in: 22 Jurimetrics (1981), S.59-82; S.57ff. D.Kaye, The Laws of Probability and the Law of the Land, in: 47 The Univ. of Chicago Law Rev. (1979), S.34-56.

1.10 Brillmayer&Kornhauser¹⁾

Auf die Kritik der beiden Autoren wird hier vor allem deshalb eingegangen, weil ihr Aufsatz in die Kategorie derjenigen Werke zu gehören scheint, die bei dem mit der Materie nicht vertrauten Leser eher Verwirrung als Aufklärung erzeugen können (auch wenn die Autoren zuweilen einen sachkundigen Eindruck erwecken).

Brillmayer&Kornhauser lehnen den Finkelstein&Fairleyschen Vorschlag²⁾ die juristische Informationsverarbeitung durch das Bayes-Theorem zu präzisieren, oder allgemeiner: die Bayes-Methode, als unwissenschaftlich ab, u.a. unter Berufung darauf, daß diese(s) bereits in der wahrscheinlichkeitstheoretischen Literatur selbst als kontrovers empfunden wird, ohne dabei zu bedenken, daß die Bayes-Methode zwar innerhalb der objektivistischen Statistik einen strittigen Punkt (eben weil die objektiven Apriori-Wahrscheinlichkeiten i.d.R. nicht verfügbar sind) darstellt, kein Problem hingegen für die subjektivistische Theorie ist³⁾

So schreiben sie (S.148,Fn.115): "It is mystifying how Finkelstein can offer such a controversial methodology without acknowledging its difficulties". Allerdings ist es eher Brillmayer&Kornhauser vorzuwerfen, daß sie die vermeintlichen Schwierigkeiten mystifizieren, erstens, weil sie zwar stets von irgendwelchen, von ihnen nicht näher präzisierten, Schwierigkeiten reden, jedoch die (eigentliche) Schwierigkeit – das Sprechen von Apriori-Wahrscheinlichkeiten in der objektivistischen Ausdrucksweise, die zu Meinungsverschiedenheiten im statistischen Lager führte, – überhaupt nicht erwähnen, geschweige denn erläutern; zweitens, weil sie weder den Bereich (d.i. die Statistik) genau abgrenzen, in welchem die Bayes-Methode kontrovers beurteilt wird, noch klar differenzieren, wer dies tut (das ist nämlich ein Teil der Statistiker, die aus den oben angegebenen Gründen sich von der praktischen Anwendung der Bayes-Methode distanzieren, jedoch

1) L.Brillmayer&L.Kornhauser, Review: Quantitative Methods and Legal Decisions, in: 46 Uni. of Chicago Law Rev. (1978), S.116-153.

2) Daß Brillmayer&Kornhauser nicht von Finkelstein&Fairleys, sondern allein von Finkelsteins Vorschlag sprechen, hat den Grund darin, daß sie sich auf das Buch von Finkelstein "Quantitative Methods in Law" (1978) beziehen, in welchem der oben diskutierte Aufsatz von Finkelstein&Fairley im 3.Kapitel wiedergegeben wird.

3) Vgl. etwa Savage, der bei der Anwendung der Bayes-Methode auch kein Indifferenzprinzip (welches von Brillmayer&Kornhauser als weiterer Vorwand gegen diese Methode benutzt wird) in Ansatz bringt.

kein Subjektivist¹⁾ drittens, weil sie nicht präzisieren, was unter der Ablehnung der Bayes-Methode seitens der "meisten Statistiker" konkret zu verstehen ist: Nicht das Bayes-Theorem, das aus den Wahrscheinlichkeitstheoretischen Axiomen folgt, wird als solches abgelehnt, sondern seine Brauchbarkeit für statistische Zwecke²⁾ und viertens, weil sie den Eindruck erwecken bzw. zu erwecken suchen - durch unklare Ausgrenzung der Schwierigkeiten, geschickte Vermengung der subjektivistischen Theorie mit dem Häufigkeitskonzept und Gleichsetzung der Problematik, die aus dem Apriori-Ansatz entsteht, mit derjenigen, die aus dem Indifferenzprinzip folgt -, daß die Schwierigkeiten primär die subjektivistische Theorie betreffen³⁾.

Von wenig Verständnis für die Wahrscheinlichkeitstheorie zeugt auch die folgende Stelle (S.136):

An important difficulty with using statistical inference about actual past historical occurrences, particularly in the judicial setting, is that the inference is not subject to the same sort of empirical verification. Introduction of Bayes's Theorem into criminal or civil trials is then curiously unscientific because the conclusions of Bayesian inference in trials are in principle untestable".

1) Brillmayer&Kornhauser weisen in ihren Fußnoten mehrmals darauf hin, daß, wie sie schreiben, "Bayesians acknowledge themselves to be in the minority" (Fn.68), wobei sie ihre Meinung immer durch ein Zitat dieses oder jenes bekannten Statistikers zu bestätigen versuchen: So die obige Behauptung durch das Zitat von Lindley, Introduction to Probability and Statistics (1965): "...most statisticians adopt a different attitude in which only frequency probabilities are admitted", ohne sich dessen offensichtlich bewußt zu sein, daß es eben "most statisticians" sind, nicht aber auch andere Wahrscheinlichkeitstheoretiker. Wegen ihrer Mißachtung der subjektivistischen Schule kann gegen Brillmayer&Kornhauser die gleiche Kritik gelten wie diejenige, die Schum gegen Cohen vorbringt (S.478): "I find it hard to believe that the author is innocent of the prevalence of the subjective interpretation of probability in decision-theoretic areas of economics, business, medicine, and psychology".

2) Eine ähnliche Kritik äußert Kaye ("Probability and the Law of the Land"), S.50f: "Contrary to what Brillmayer and Kornhauser seem to imply (Fn.54: Brillmayer and Kornhauser seem impressed with the fact that "Bayesians" have not yet written many elementary textbooks...), the objection to using subjective prior probabilities does not deny that Bayes's Theorem is mathematically correct" (in Fn.57: "Bayes's Theorem is part and parcel of the theory of probability presentend in every elementary text on the subject. It is all but implicit in the very definition of conditional probability. It can be used with either objectively or subjectively determined probabilities").

3) Ihrer Ablehnung der Bayes-Methode verleihen die Autoren auch auf eine rhetorische Weise Ausdruck: So sprechen sie vielerorts von "vielen", den "meisten", "weltbekannten", "prominenten" und "eminenter" "Anti-Bayesianern" (wobei meistens ältere Autoren wie Bernoulli (1654-1705), Keynes (1921), Nagel (1939) zitiert werden) und nur von einer "Minderheit" der Bayesianer, die "sich dessen auch bewußt sei" (vgl. S.135 Fn.68, S.141 Fn.91, S.148f Fn.115).

Etwas ausführlicher zur Auffassung, daß Brillmayer&Kornhauser offensichtlich die Problematik des Bayes-Ansatzes völlig mißverstanden haben, äußert Kaye, *ibid.*, S.51, Fn.57.

Die oben zitierte Behauptung ist vor allem aus zwei Gründen abzulehnen: Erstens wird hier postuliert, daß bei der Beurteilung juristischer (historischer) Ereignisse nur die statistische Betrachtungsweise in Frage käme und dementsprechend die Anwendung des Bayes-Theorems nicht korrekt eingeschränkt¹⁾ und zweitens wird dieses aufgrund von etwas abgelehnt, das es eigentlich nicht leisten soll: Das Bayes-Theorem ist keine Verifikations- bzw. Falsifikations-, sondern eine Operationsmethode (d.h. eine Rechenmethode), die gegenüber dem intuitiven (und von Brilmayer&Kornhauser verteidigten) Ansatz den Vorteil besitzt, auf konsistente, konsequente und übersichtliche Weise probabilistische Annahmen miteinander zu verknüpfen.

Die Bayes-Methode als "unwissenschaftlich" abzulehnen, wie die Autoren es tun, weil sie zu keinen "empirisch verifizierbaren" Ergebnissen führe, bedeutet außerdem, an dem radikalen Verifikationspositivismus (der in der Wahrscheinlichkeitstheorie längst als bereits überwunden gilt) festzuhalten, der konsequent durchgeführt zur Ablehnung v.a. der objektiven Wahrscheinlichkeitstheorie führen würde.

1) Dies ist um so verwunderlicher als die Autoren an einer späteren Stelle die subjektivistische Interpretation in Erwägung ziehen (vgl. S.139f) und außerdem auch Finkelstein wegen seines Versuchs, den subjektivistischen Wahrscheinlichkeitsbegriff frequentistisch umzu-deuten, angreifen (vgl.S.140).

(Der subjektivistische Wahrscheinlichkeitsbegriff erfährt allerdings bei ihnen eine quasi frequentistische Deutung, vgl. S.140f: "A more accurate statement of the Bayesian view is that if we have a degree of belief of .51, this figure is the equivalent of saying that we think that out of one hundred cases we would be right fifty-one times, not that we would be correct fifty-one times". Entgegen der Auffassung von Brilmayer&Kornhauser scheint es mir, daß sie nicht den Begriff der subjektiven Wahrscheinlichkeit 'explizieren', sondern den Häufigkeitsbegriff subjektivistisch uminterpretieren).

Daß die Autoren die Anwendung des Bayes-Theorems statistisch "einschränken", bestätigt auch ihre Äußerung, daß jenes nicht auf individuelle Fälle anwendbar sei, vgl.S.150: "The objection that statistics appears to deny people treatment as individuals could possibly arise even when the characteristic in question is under the individual's control. One wonders whether some of the objections to use of Bayesian methods do not stem from the sense that the trier-of-fact should be concerned with accuracy in the particular case, and not with an overall greater record of accuracy".

Und schließlich muß man sich darüber klar sein, was man eigentlich "verifizieren" will. Sind es juristische (also vergangene) Ereignisse, dann geht es beim von Brilmayer&Kornhauser angesprochenen Verifizieren in Bezug auf die Bayes-Methode offensichtlich darum, ob sie bestehen oder nicht bestehen und nicht darum, ob sie in diesem oder jenem "objektiven" Grad bestehen¹⁾. Diese Art des Verifizierens kann aber nicht nur die probabilistische, sondern auch die intuitive Beurteilung nicht liefern. Hingegen vermag man wie bereits erwähnt, durch probabilistische Methoden das Überzeugungsausmaß des Individuums im subjektivistischen Rahmen wohl zu verifizieren²⁾.

Außer daß sie den Bayes-Ansatz mißverstanden haben, verwechseln Brilmayer und Kornhauser darüber hinaus oft methodologische Fragestellungen mit wahrscheinlichkeitstheoretischen, wie die folgende Stelle zeigt (S.145):

"The legal system appears to contain separate requirements that some certain amount of specific evidence be adduced before a prima facie case is established. If the plaintiff in a civil case produces a minimal amount of evidence, a motion for a directed verdict may be granted even if the defendant has as yet

1) Zu diesem Punkt vgl. S.68f dieser Arbeit. Brilmayer&Kornhauser meinen offensichtlich das "Verifizieren" im letzteren Sinn; vgl. Brilmayer&Kornhauser, S.135f.

2) Aus diesem Grund ist auch die folgende Behauptung Brilmayer&Kornhauser nicht korrekt (S.147, Fn.111): "Since frequency statements have empirical content they can, in principle, be empirically tested; it is the controversial subjective "degree of belief" interpretation that cannot". Faßt man "empirisch prüfbar" in Sinne "empirisch verifizierbar bzw. falsifizierbar" auf, dann verhält es sich aus wahrscheinlichkeitstheoretischer Sicht gerade umgekehrt: Es sind nämlich objektive Wahrscheinlichkeitsaussagen, die, wie bekannt, weder verifizierbar noch falsifizierbar sind. Das war auch einer der beiden Hauptgründe, die z.B. de Finetti dazu bewogen haben, die objektivistische Wahrscheinlichkeitstheorie abzulehnen (vgl. S.45f dieser Arbeit). Die subjektiven Wahrscheinlichkeitsurteile (in Form von Glaubensannahmen) sind hingegen immer entscheidbar (feststellbar), d.h. man kann immer entscheiden, ob eine Person in diesem oder jenem Grad an etwas glaubt (zu diesem Punkt vgl. S.45 dieser Arbeit). (Die subjektivistischen Wahrscheinlichkeitsaussagen sind nichts anderes als eine explizite und rationale Darstellung der intuitiven Überzeugung eines Individuums).

introduced no evidence at all. This procedure seems to indicate that the legal system has methods of measuring the total quantity of evidence quite separate from and in addition to its methods of comparing the relative weight of the evidence on either side. Therefore, a probability figure that reflects the extent to which a given body of evidence supports an inference, but not the sufficiency of the body of evidence itself, cannot adequately capture all of the elements of legal proof".

Das letztere ist zwar richtig, nur muß man auch hier klar sehen, was die Rechtsprechung in der Praxis erwartet/fordert. Sie fordert, daß für die Beurteilung/Lösung ein hinreichendes Tatsachenmaterial vorliegt. Diese Forderung des hinreichenden Datums kann natürlich nicht innerhalb der Wahrscheinlichkeitstheorie formuliert werden. Viel mehr muß sie aus methodologischen bzw. juristischen Gründen zusätzlich als Voraussetzung der Anwendung der Wahrscheinlichkeitstheorie auf die Rechtspraxis aufgestellt werden¹⁾

Auf drei weitere Einwände seitens Brilmayer&Kornhauser sei hier kurz eingegangen.

Auf S.145 folgt die Wiederholung der von Cohen gegen die Wahrscheinlichkeitstheorie vorgebrachten Kritik (die sog. "Cohensche erste Paradoxie"), die darin besteht, daß das Multiplizieren mehrerer Wahrscheinlichkeiten miteinander eine geringere Wahrscheinlichkeit für das gemeinsame Vorliegen aller Beweismittel bewirkt als eine solche für jedes einzelne, was Cohen und Brilmayer&Kornhauser zufolge mit der juristischen Intuition unverträglich ist. Da diese Schwierigkeit im Rahmen meiner Kritik an Cohen gelöst wird, soll auf sie hier nicht weiter eingegangen werden²⁾

1) Zu diesem Punkt, d.h. daß das Bayes-Modell abzulehnen wäre, weil es keine methodologischen Beweiskriterien formuliert vgl. auch Brilmayer&Kornhauser (S.142): "One rational strategy that does not fit the Bayesian model is a court's finding of "insufficient evidence", that is, that a party simply has not adduced sufficient evidence to warrant disturbing the status quo. (Fn.94: In a civil case such a finding means that the plaintiff has not satisfied his burden of proof; in a criminal case it means that the state has failed to establish guilt beyond a reasonable doubt)".

Zur Formulierung und Berücksichtigung dieser Forderung (des. sog. "hinreichenden Datums") bei der probabilistischen Beurteilung eines juristischen Falls vgl. S.214 dieser Arbeit.

2) Vgl. Brilmayer&Kornhauser (S.145): "Bayesian principles suggest the combined probability on the issues is significantly less than one half" und weiter (Fn.103): "This example is taken from L.Cohen, The Probable and the Provable 58-59 (1977) an insightful and challenging account of the difficulties in Bayesian proof schemes which attempts a non-Bayesian reformulation". Erstens ist das obige Prinzip kein Bayessches Prinzip, sondern das Multiplikationsprinzip und zweitens hat Cohen das Bayes-Theorem mit keinem Wort erwähnt, geschweige denn "the difficulties in Bayesian proof schemes" gezeigt.

Der zweite Einwand (S.146) betrifft die sog. 'Anfechtung' des Tatsachenmaterials bei den Zeugenaussagen. Vgl.: "Consider a case that turns on whether an automobile driver ran a red light. The pedestrian who is suing him for negligence did not see whether the light had changed and must rely upon the testimony of a single witness who says that it was still red. Unfortunately, cross-examination shows that the witness is somewhat forgetful, so that the jury grants him a credibility of only one half. Assessing the probability of negligence based on the evidence just described proves to be rather troublesome".

Wir können den obigen Fall in der Tat mit dem Bayes-Theorem allein nicht lösen, jedoch nicht in dem von Brillmayer&Kornhauser gemeinten Sinn. Denn Fälle, die von der Unsicherheit nicht nur in bezug auf die Hypothese, sondern auch auf die Tatsachen, die der Hypothese zugrunde liegen, ausgehen, sind entweder mit der erweiterten Bayes-Formel (vgl. S.34 dieser Arbeit) oder mit Hilfe der Jeffrey-Formel (deren Bestandteil das Bayes-Theorem bildet) korrekt lösbar.

Der dritte Einwand¹⁾ besteht darin, daß Brillmayer&Kornhauser meinen, daß es der juristischen Intuition zuwiderläuft, wenn dieselbe Information sowohl die Hypothese wie auch ihre Negation wahrscheinlich macht²⁾ vgl. S.146:

"The impeachment of the witness does not tell us that what the witness is saying is false; it merely negates the inference that it is true. The jury would be quite rational in describing its attitudes after hearing this evidence by saying that it was half convinced that the light was red, but had heard no evidence whatsoever tending to show that the light was green. This description of the probative value of the evidence is quite in line with the law's attitude towards impeachment evidence; the evidence could not be offered directly to show that the light was green, since it does not tend to show that the light was green. There is a difference between refuting a proof that is given and proving that the opposite is true".

Falls die juristische Intuition tatsächlich so funktionieren würde, wie Brillmayer&Kornhauser und Cohen behaupten, dann könnten wir mindestens drei Gründe anführen, warum eine solche Intuition nicht rational wäre:

1. Man gerät in die auf S.70f beschriebene Schwierigkeit. Zur Veranschaulichung knüpfen wir an das obige Beispiel an:
Sei H die Hypothese, daß die Ampel grün zeigte und $\neg H$, daß sie rot war. Ferner sei A die Zeugenaussage, die mit an Sicherheit grenzender Wahrscheinlichkeit (etwa 0.9 für die beurteilende Person) beweist, daß die Ampel grün zeigte, d.h. $P(H/A) = 0.9$, und B die Zeugenaussage, die mit an Sicherheit grenzender Wahrscheinlichkeit beweist, daß das Licht rot war, d.h. $P(\neg H/B) = 0.9$ (auch ein falsches Zeugnis, sei es, daß der Zeuge lügt, sei es, daß der Zeuge zwar subjektiv, jedoch nicht objektiv, die Wahrheit sagt, kann sehr glaubwürdig erscheinen).

1) Auch hier ist der Einfluß von Cohen auf Brillmayer&Kornhauser unverkennbar.

2) Zu diesem Punkt vgl. S.70ff dieser Arbeit, wo ich mich mit dieser Problematik ausführlich befasse.

Aus wahrscheinlichkeitstheoretischer Sicht können wir in einem solchen Fall ein plausibles Ergebnis erreichen. Z.B. mit $P(H) = 0.5$, $P(A/H) = P(B/\neg H) = 0.9$ und $P(A/\neg H) = P(B/H) = 0.1$ (die Zeugen sind bestimmt in einem solchen Fall voneinander unabhängig) erhalten wir aufgrund dieser beiden Zeugenaussagen $P(H/A \wedge B) = P(\neg H/A \wedge B) = 0.5$. Also ist die Wahrscheinlichkeit der Hypothese aufgrund der beiden gleichermaßen glaubwürdigen Zeugenaussagen, die jeweils zugunsten des Gegenteils sprechen, die gleiche wie vorher.

Wie können wir aber vorgehen, wenn unsere Intuition so arbeitet, wie Brilmayer & Kornhauser den Juristen unterstellen? Offensichtlich müssen wir dann entweder willkürlich zugunsten einer der beiden Hypothesen (die ja beide mit an Sicherheit grenzender Wahrscheinlichkeit bewiesen sind) entscheiden, oder hoffen, daß so etwas vor Gericht nicht vorkommt oder zu irgendeiner sonstigen irrationalen Strategie greifen.

2. Man gerät außerdem auch in Schwierigkeit mit einigen juristischen Grundsätzen, etwa mit demjenigen der "überwiegenden Wahrscheinlichkeit". Sei $P(H) = 0.6$. Ferner nehmen wir an, daß die beiden Zeugenaussagen, die jeweils zugunsten des Gegenteils sprechen, dem Richter für die Beurteilung des Falls ausreichen; allerdings beträgt für ihn $P(H/A) = 0.3$ und $P(\neg H/B) = 0.2$. Nach dem Grundsatz der überwiegenden Wahrscheinlichkeit muß er zugunsten des Sachverhalts entscheiden, obwohl seine Wahrscheinlichkeit relativ auf die zu seinen Gunsten sprechende Zeugenaussage, unvoreingenommen gesehen, in der Tat sehr gering ist. Im wahrscheinlichkeitstheoretischen Fall können wir wiederum ein plausibles Ergebnis erreichen: Sei z.B. $P(A/H) = 0.3$, $P(A/\neg H) = 0.9$, $P(B/H) = 0.8$, $P(B/\neg H) = 0.3$. Wir erhalten somit $P(H/A \wedge B) = 0.57$ und $P(\neg H/A \wedge B) = 0.43$. Auch hier überwiegt (aufgrund des Gesamtdatums) zwar die Hypothese; dieses Ergebnis können wir aber besser rechtfertigen: Denn indem die beiden Zeugen jeweils zugunsten des Gegenteils sprechen, es aber weniger wahrscheinlich als vorher machen (offensichtlich handelt es sich bei diesen Zeugen um sehr unglaubliche, z.B. sich widersprechende Personen), lassen sie Zweifel an ihren jeweiligen Schilderungen des Falls aufkommen. D.h. das unglaubliche Verhalten der Zeugen A und B veranlaßt den Richter, jedesmal viel stärker an das Gegenteil dessen zu glauben, was der jeweilige Zeuge berichtet, was seinen Ausdruck darin findet, daß die Zeugenaussage zugunsten der Hypothese diese nur zu 0.3, die Negation der Hypothese hingegen zu 0.7 wahrscheinlich macht und entsprechend die Zeugenaussage zugunsten der Negation der Hypothese diese nur zu 0.2, die Hypothese hingegen zu 0.8 wahrscheinlich macht. Da die erste Zeugenaussage dem Richter etwas unglaublicher erscheint als die zweite, ist die Apriori-Wahrscheinlichkeit der Hypothese aufgrund dieser beiden Zeugenaussagen kleiner als die Apriori-Wahrscheinlichkeit.

Man gerät auch mit dem Grundsatz "mit an Sicherheit grenzender Wahrscheinlichkeit" in Schwierigkeiten. Sei z.B. für den Richter in einem konkreten Fall $p = 0.99$ der äquivalente numerischer Ausdruck für den qualitativen Begriff "mit an Sicherheit grenzender Wahrscheinlichkeit". Ferner seien $P(H/A) = 0.99$ und $P(\neg H/B) = 0.98$. D.h. trotz einer sehr hohen Wahrscheinlichkeit für die Negation der Hypothese muß H von einem 'Brilmayer-Kornhauser'-Richter als "wahr" akzeptiert werden¹⁾

1) Sei $P(H) = 0.5$. Ferner seien $P(A/H) = 0.99$, $P(A/\neg H) = 0.01$, $P(B/H) = 0.02$ und $P(B/\neg H) = 0.98$. Wir erhalten somit $P(H/A) = 0.99$ und $P(\neg H/B) = 0.98$, was zu dem oben geschilderten Dilemma führt. Ein rationaler Richter hingegen erhält aufgrund der beiden Informa-

3. Brillmayer&Kornhausers Behauptung, daß der juristische Sachverhalt nicht im Lichte aller, sondern nur in dem der für ihn günstigen Informationen beurteilt wird, verletzt außerdem eine der wichtigsten methodologischen Rationalitätsforderungen (die für die rationale Informationsverarbeitung wichtig ist), nämlich, daß ein Ereignis aufgrund aller für ihn relevanten (d.h. seine Glaubwürdigkeit verändernden: bestätigenden oder bezweifelnden) Informationen beurteilt werden muß.

Daß man die von Brillmayer&Kornhauser postulierte Vorgehensweise etwa im Alltag als irrational empfinden würde, veranschaulicht das folgende Beispiel: Ein Geschäftsmann berücksichtigt beim Kauf bestimmter Aktien nur diejenigen Informationen, die ihm (gleich mit welcher Wahrscheinlichkeit) bestätigen, daß diese Aktien in ihrem Wert steigen; in bezug auf die Negation der Hypothese (d.h. daß die Aktien nicht steigen), betrachte er sie als "nicht informativ". Angenommen, zufällig bestätigen in letzter Zeit alle Informationen die Hypothese über das Steigen der Aktien, machen sie aber insgesamt nur gering wahrscheinlich (weil diese Informationen vielleicht nicht besonders glaubwürdig sind)¹⁾. Falls dieser Geschäftsmann nun sich entschließt, die Aktien zu kaufen, so würden wir seine Entscheidung für höchst bedenklich halten, da vielmehr zu erwarten ist, daß die Aktien (relativ auf die gesamte Information, die zwar die Hypothese bestätigt, deren Wahrscheinlichkeit aufgrund dieser Information aber noch immer gering ist) nicht steigen werden. Dieser Geschäftsmann würde bei dieser Strategie bald Bankrott gehen, womit ihm auf empfindliche Weise gezeigt wäre, daß er unvernünftig gehandelt hat. (Leider fehlen Juristen derartige beweiskräftige "Überzeugungsmittel", der einzige, der bei Brillmayer&Kornhauser Strategie "Bankrott gehen" kann, ist der Angeklagte).

tionen $P(H/A \wedge B) = 0.67$ und $P(\neg H/A \wedge B) = 0.33$. D.h. werden bei der Beurteilung der Hypothese Informationen berücksichtigt, die sie sowohl wahrscheinlicher als auch unwahrscheinlicher machen, so wird die Hypothese nie einen dermaßen hohen Wahrscheinlichkeitsgrad erreichen, wie man für ihn im Lichte der für sie günstigen Informationen herleiten würde.

1) Zur Illustration: Sei $P(H) = 0.1$, also $P(\neg H) = 0.9$. Weiter seien E, E' Informationen, die H bestätigen, wobei gilt: $P(H/E) = P(H/E') = 0.2$. Wir erhalten: $P(H) = 0.1 < P(H/E) = 0.2 + P(\neg H/E) = 0.8 < P(\neg H) = 0.9$ und $P(H) = 0.1 < P(H/E') = 0.2 + P(\neg H/E') = 0.8 < P(\neg H) = 0.9$. Ferner folgt (mit der Bayes-Formel): $P(H/E \wedge E') = 0.36 < P(\neg H/E \wedge E') = 0.64$, d.h. die Wahrscheinlichkeit der Hypothese aufgrund der beiden sie bestätigenden Informationen ist noch immer kleiner als eine solche für die Negation der Hypothese aufgrund dieser Informationen. (Bei der o. Berechnung setzen wir voraus, daß E und E' voneinander unabhängig sind).

1.11. Diskussion des Wahrscheinlichkeitsbegriffs in Deutschland

Wie bereits erwähnt¹⁾ trägt die Diskussion des Wahrscheinlichkeitsbegriffs in Deutschland einen informellen Charakter: Man ist nicht am methodologischen Aspekt der Wahrscheinlichkeitstheorie, sondern an der Klärung der Existenzberechtigung der Wahrscheinlichkeitsurteile für die Rechtspraxis interessiert. Auffällig und erstaunlich ist die Naivität der meisten deutschen Autoren im Umgang mit den Ausdrücken "objektive" und "subjektive" Wahrscheinlichkeit, was den Verdacht erweckt, daß jene oft nicht verstehen, wovon sie reden²⁾

Da die Auseinandersetzung mit den deutschen Beiträgen zur Diskussion des Wahrscheinlichkeitsbegriffs bereits Gegenstand einer Arbeit ist³⁾ möchte ich den Leser mit Wiederholungen nicht strapazieren, zumal ich mich sowohl der Kritik dieses Autors an dem unreflektierten Gebrauch des objektiven Wahrscheinlichkeitsbegriffs in der deutschen juristischen Literatur wie auch seiner Auffassung, daß es sich bei dem juristischen Wahrscheinlichkeitsbegriff um einen subjektiven handelt, völlig anschließe⁴⁾

1) Vgl. S. 63ff dieser Arbeit; der Leser findet dort auch die Literaturangaben.

2) *ibid.*, auch S. 67f dieser Arbeit.

3) E. L. Nell, *Wahrscheinlichkeitsurteile in juristischen Entscheidungen* (Dissertation), 1983. Nell vertritt, wie bereits gesagt (S. 62, Fn. 1 dieser Arbeit) den subjektivistischen Standpunkt.

4) Zwar tendiert auch Evers (A. Evers, *Begriff und Bedeutung der Wahrscheinlichkeit für die richterliche Beweiswürdigung* (Dissertation), 1979) in die subjektivistische Richtung, mißversteht allerdings diese Auffassung völlig, wenn er davon ausgeht, daß die Metrisierung von Wahrscheinlichkeitsaussagen im subjektivistischen Rahmen nur dann möglich ist, wenn irgendwelche allgemeingültigen Erfahrungssätze vorliegen (vgl. Evers, S. 106, 148), wobei er als Kriterium der Allgemeingültigkeit solcher Erfahrungssätze ihre Wiederholbarkeit ansieht; vgl. S. 160: es "bleibt dabei doch richtig, daß die Wiederholung oder Häufigkeit einzelner Vorgänge ein entscheidendes Kriterium zur Bestimmung des nach der subjektivistischen Wahrscheinlichkeitstheorie zu ermittelnden Grades von Wahrscheinlichkeit ist. Dies ergibt sich daraus, daß der nach der subjektiven Wahrscheinlichkeitstheorie erreichbare Grad von Wahrscheinlichkeit entscheidend davon beeinflusst wird, als wie stark eine bestimmte Erfahrung allgemein anzusehen ist".

Dies steht jedoch im Widerspruch zur herkömmlichen Auffassung der Subjektivisten, wonach die Metrisierung der qualitativen Wahrscheinlichkeitsurteile nicht davon abhängt, wie allgemein bzw. allgemeingültig bzw. wiederholbar die Erfahrung ist, sondern einzig und allein davon, wie das Subjekt im Lichte seiner eigenen persönlichen Erfahrung (diese muß nicht allgemein sein) das Ereignis beurteilt.

Evers bringt anscheinend zwei Dinge durcheinander: zum einen die Metrisierung qualitativer Wahrscheinlichkeitsurteile im subjektivistischen Rahmen und zum anderen das Erreichen der intersubjektiven Übereinstimmung aufgrund gemeinsamer Erfahrungen. Hängt das erste nur von der persönlichen Erwartungshaltung des Subjekts dem Eintreten eines Ereignisses gegenüber ab, so ist das zweite in der Tat davon abhängig, wie allgemein Erfahrungen sind und wie gleichmäßig diese von verschiedenen Personen bewertet werden.

Vgl. auch Evers, S. 176: "Zunehmende empirische Information führt also zur Begründung steigender subjektiver Wahrscheinlichkeit" (?), der gleiche Fehler also wie bei Kegel (vgl. S. 99, Fn. 1 dieser Arbeit).

Zusammenfassung und Ausblick¹⁾

In diesem Paragraphen haben wir die wesentlichen Ergebnisse aus der Diskussion des Wahrscheinlichkeitsbegriffs in der juristischen Literatur von 1961 bis ca. 1971 geschildert. Wir haben dabei sowohl die vorgeschlagenen Methoden wie auch die Kritik an diesen Methoden betrachtet. Unser Anliegen war es, den Leser v.a. mit folgenden Aspekten dieser Diskussion bekanntzumachen:

1. Den wichtigsten Techniken für eine Bestimmung der Sachverhalts-hypothese. Diese sind: die von Kingston eingeführte und von Cullison vereinfacht dargestellte Binomialformel für die Bestimmung der Identitäts- bzw. Duplizierungswahrscheinlichkeit, wobei uns einzig und allein statistische Daten zur Verfügung stehen; und die von Finkelstein und Fairley vorgeschlagene Methode, mittels des Bayes-Theorems die Wahrscheinlichkeit der Beweismittelannahme relativ zu den zur Verfügung stehenden Informationen zu bestimmen, wobei das nichtstatistische und das statistische Datum in rationaler Weise miteinander verknüpft werden kann. Mittels dieser Methode können wir somit auch unsere intuitiv-subjektiven Schätzungen in konsistenter Weise zum Gesamturteil verarbeiten²⁾.

Diese beiden Methoden sind übrigens auch für den aktuellen Stand der Diskussion von Bedeutung. Für die Zeit nach 1971 können wir lediglich zwei Verbesserungsvorschläge der von Kingston vorgeschlagenen Methode anführen: diejenige von Smith&Charrow (1975) und diejenige von Stripinis (1981)³⁾.

2. Anderen (allerdings nicht korrekten oder nicht adäquaten) Techniken. Diese sind: die von Schreiber vorgeschlagene Methode für die Feststellung der Wahrscheinlichkeit für den Beweiswert mehrerer hinreichender Beweismittel; und die von Ekelöf eingeführte Formel für die Bestimmung des Gesamtwertes zweier Beweismittel.

3. Den kritischen Auseinandersetzungen sowohl positiver wie auch negativer Art. Während Cullison programmatisch für die erste steht - er bejaht die Bedeutung der Wahrscheinlichkeitstheorie für die Rechtspraxis, kritisiert aber den schlampigen Gebrauch solcher Me-

1) Eine ausführliche Kritik an einzelnen Ansätzen findet der Leser in den jeweiligen Abschnitten.

2) Meine Kritik an Finkelstein&Fairleys Ansatz bestand v.a. darin, daß man die Bayes-Formel nicht nur - wie die beiden das ausdrücklich betonten - auf die Beweismittelannahme, sondern auch auf die juristische Sachverhaltshypothese und nicht nur auf das statistische Datum, sondern allgemein auf jedes Datum anwenden kann; vgl. S.125 dieser Arbeit.

3) Vgl. aber auch den ziemlich technischen Aufsatz von J.W.Osterburg, T.Parthasarathy, T.E.S.Raghavan und S.L.Sclove, Development of a Mathematical Formula for the Calculation of Fingerprint Probabilities Based on Individual Characteristics, in: 72 Journal of American Stat.Ass. (1977), S.772-778.

Weitere Aufsätze, die sich mit der Anwendung der Wahrscheinlichkeitstheorie auf die Rechtspraxis befassen, sind: D.V.Lindley, Probabilities and the Law, in: Utility, Probability and Human Decision Making (1975), S.223-232; W.B.Fairley, Probabilistic Analysis of Identification Evidence, in: ibid, S.233-256.

thoden durch Juristen, die diese Methoden nicht beherrschen -, vertritt Tribe im Gegensatz dazu die Auffassung, daß die Wahrscheinlichkeitstheorie wesentlich mehr schadet als nutzt, weil sie nicht nur die Juristen dazu motiviert, sie zu mißbrauchen, sondern weil sie, seiner Meinung nach, auch die traditionellen juristischen Werte zerstört. Demgegenüber lehnen Brilmayer und Kornhauser die Wahrscheinlichkeitstheorie zum größten Teil aus falschverstandenen konzeptionellen und epistemologischen Gründen ab.

Außerdem haben wir auch die Arbeiten derjenigen Autoren diskutiert, die zwar keine ausgefeilten Techniken bieten, jedoch für die Wahrscheinlichkeitstheoretische Beurteilung der Rechtspraxis eintreten. Dies sind die Arbeiten Balls¹⁾ und Kegels.

Im Unterschied zu diesem Paragraphen, der zum größten Teil theoretischer Art war, wollen wir uns im nächsten mit dem empirischen Aspekt der Anwendung der Wahrscheinlichkeitstheorie auf die Rechtspraxis beschäftigen. Wir wollen anhand der Ergebnisse mehrerer empirischer Untersuchungen die Frage beantworten, inwiefern die Wahrscheinlichkeitstheorie auch praktisch für Gerichtsverfahren relevant sein kann.

1) Ball benutzt, wie gezeigt wurde, zum ersten Mal die Boolesche Formel. Zur Kritik an dieser Formel siehe S.86, Fn.1, 109f dieser Arbeit.

§2. Empirische Untersuchungen

In diesem Paragraphen werden Ergebnisse einer Reihe empirischer Untersuchungen geschildert, die sich mit der Frage befassen, inwiefern die Wahrscheinlichkeitstheorie für die Rechtspraxis relevant ist bzw. sein kann.

Anhand dieser Ergebnisse wird gezeigt, daß manche übliche Vorstellungen über die Urteilsfähigkeit von mit der Wahrscheinlichkeitstheorie nicht vertrauten Personen revidiert werden müssen. Diese sind:

1. daß Laien und Juristen mit Wahrscheinlichkeiten nicht umgehen können (wie wir wissen, war das einer der Gründe für die ablehnende Haltung Tribes gegenüber einer Anwendung der Wahrscheinlichkeitstheorie in der Rechtspraxis);
2. daß die intuitive Verarbeitung der Information zum Gesamtbild dem Bayesianischen Ansatz widerspricht (Tribes, aber auch Cohen);
3. daß die "an Sicherheit grenzende Wahrscheinlichkeit" in den Augen derjenigen, die juristische Fälle beurteilen, "nur erkenntnistheoretische Zweifel" (wobei diese nicht numerisch ausdrückbar sind) repräsentiert (d.h. die Auffassung der theoretisierenden Rechtsprechung, daß es sich bei dem obigen Ausdruck um einen kontexteliminierbaren Begriff handelt¹⁾ wird empirisch nicht bestätigt. Ebenso wenig wird aber auch die Auffassung derjenigen bestätigt, welche die "an Sicherheit grenzende Wahrscheinlichkeit" durch die Zahl 0.99... ausdrücken wollen²⁾);
4. daß die Anwendung der Wahrscheinlichkeitstheorie auf die Rechtspraxis alles andere als sinnvoll ist (etwa Tribes). Gerade die explizite Anwendung solcher Methoden und auch die Quantifizierung von subjektiven Urteilen seitens der Beurteilenden kann uns nicht nur zur konsistenten Informationsverarbeitung im Rechtsprozeß führen, sondern auch Aufschluß über die Gründe für die unterschiedliche Behandlung identischer Fälle (und für die unterschiedlichen Schlußfolgerungen) seitens der Juristen geben.

2.1. Simon und Mahan³⁾

In ihrer empirischen Untersuchung kommen Simon & Mahan zu folgenden wichtigen Ergebnissen:

1. Laien können ohne weiteres ihre subjektiven Urteile über die Täterschaft des Angeklagten durch numerische Wahrscheinlichkeitswerte repräsentieren;
2. die numerische Darstellung des Urteilsspruchs führt in der Regel zu einer sorgfältigeren Abwägung der Beweismittel und zu einer günstigeren Einstellung gegenüber dem Angeklagten.
3. Der durchschnittliche Wert für den Ausdruck "mit an Sicherheit grenzender Wahrscheinlichkeit" liegt viel niedriger und der Unterschied zwischen diesem Wert und dem durchschnittlichen Wert für den Ausdruck "mit überwiegender Wahrscheinlichkeit" ist viel geringer als man gewöhnlich annimmt.

Im folgenden wollen wir auf die Untersuchung der beiden Wissenschaftlerinnen näher eingehen.

Ziel der Untersuchung war es, neue Möglichkeiten zu erforschen, die im Straf- und Zivilgericht benutzten Beweislastregeln präziser

1) Vgl. S. 58 Fn. 1 dieser Arbeit.

2) *ibid.*

3) R.J. Simon & L. Mahan, Quantifying Burdens of Proof, in: 5 Law and Society Review (1971), S. 319-330. Mit einer ähnlichen Problematik befaßt sich auch der Aufsatz von Simon, "Beyond a Reasonable Doubt" - An Experimental Attempt at Quantification (1970), auf den hier nicht näher eingegangen wird.

und objektiver zu definieren¹⁾

Ihre Versuchsgruppe bestand aus 69 Geschworenen²⁾ und 88 Soziologiestudenten. Daraus wurden insgesamt 10 "Geschworenengerichte" und 14 "Studentengerichte" gebildet. Die Geschworenengerichte bat man, sich die redigierte Aufzeichnung eines Gerichtsprozesses, in dem die Anklage auf Totschlag lautete, anzuhören während man vor den Studentengerichten den gleichen Prozeß vorspielte. (Ein weiterer Unterschied bestand darin, daß die Geschworenen in einem Gerichtssaal, die Studenten in einem Übungsraum versammelt waren). Nach der Anhörung des Falls forderte man die Teilnehmer an der Untersuchung auf, 30 min. über ihre Entscheidungen zu beraten. Unmittelbar vor der Beratung und gleich danach bat man sie, ihre individuellen Entscheidungen auf einem Fragebogen anzugeben. Die Teilnehmer wurden dabei in zwei Gruppen geteilt: Die erste Gruppe mußte ihre individuellen Entscheidungen zuerst in der traditionellen Weise als "schuldig"/"unschuldig"-Urteile angeben, sodann sie in Wahrscheinlichkeitsurteile übertragen, während die zweite umgekehrt vorzugehen hatte (vgl. die beiden Fragebogen; Fig.1 und 2)³⁾

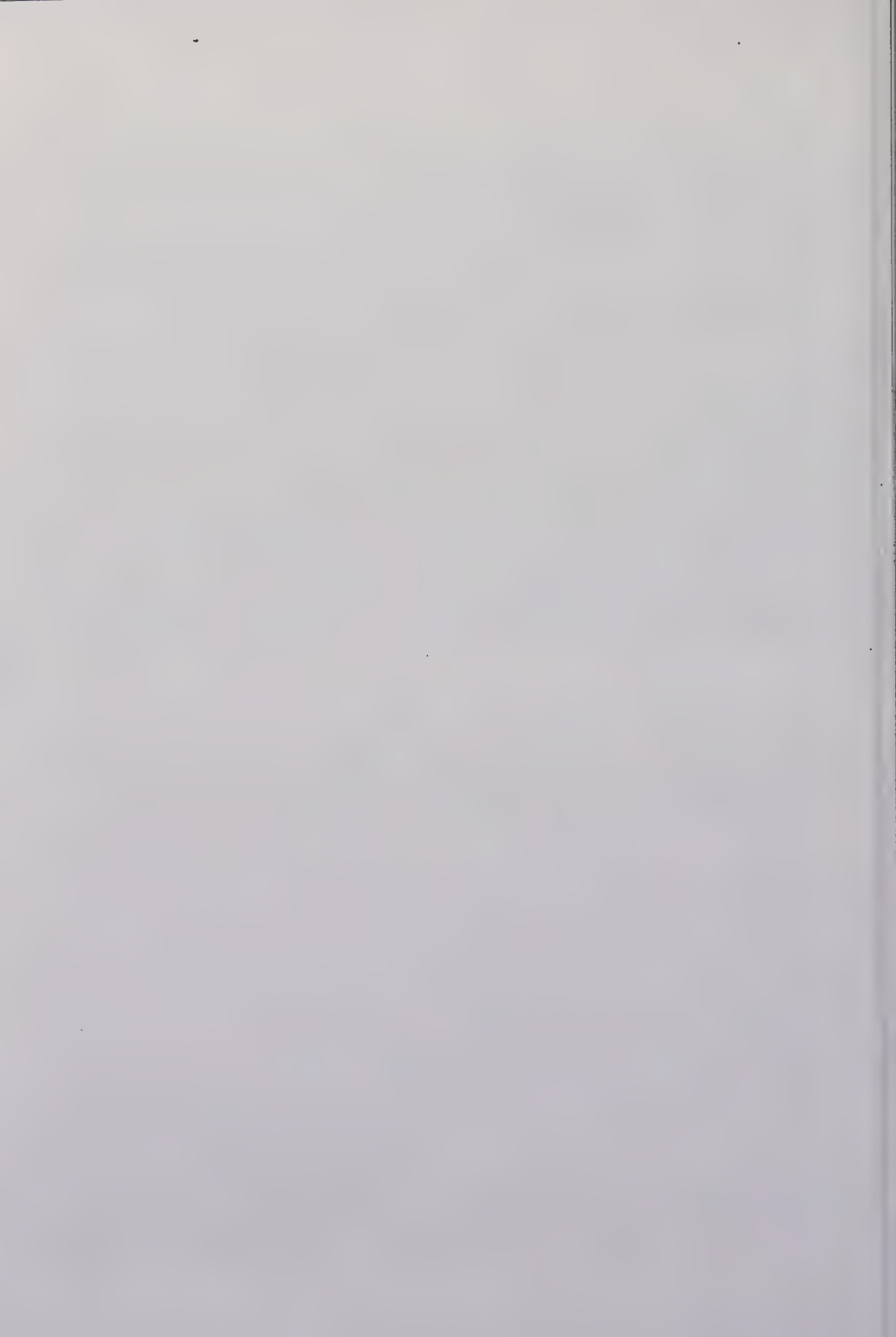
1) Simon&Mahan befassen sich mit amerikanischen Gerichten. Hier gelten folgende Beweiskriterien: "mit an Sicherheit grenzender Wahrscheinlichkeit" ("beyond a reasonable doubt"; in dieser Hinsicht ähnelt die amerikanische Gerichtspraxis der deutschen) im Strafrecht und "mit überwiegender Wahrscheinlichkeit" ("by preponderance of the evidence") im Zivilrecht (was diese Regel angeht, so wird sie in Deutschland zwar eifrig diskutiert, allerdings von den meisten Gerichten nicht akzeptiert).

Simon&Mahan geben an, was man in der amerikanischen Rechtspraxis unter diesen beiden Beweiskriterien versteht; vgl. S.330:

"Reasonable doubt is one a reasonable person has after carefully weighing all the testimony and is one a reasonable person would act or decline to act upon. It is not a capricious doubt or a fanciful doubt or a doubt arising in anyone's mind because of any sympathy for the defendant. It is in essence what the words obviously mean - a reasonable doubt. A reasonable doubt may arise not only from the evidence produced but also from a lack of evidence"; und weiter: "To "establish by a preponderance of the evidence" means to prove that something is more likely so then not so. In other words, a preponderance of the evidence in the case means such evidence as, when considered and compared with that opposed to it, has more convincing force, and produces in your minds belief that what is sought to be proved is more likely true than not true".

2) Die Geschworenen wurden mit Hilfe des Chief Judges of the Campaign County Court ausgesucht und standen zur Zeit der Untersuchung bereits in ihrem Dienst (Geschworene werden auch in Amerika bekanntlich aus allen Bevölkerungsschichten ausgewählt).

3) Alle nachfolgenden Fragebogen und Tabellen sind dem obengenannten Aufsatz von Simon&Mahan photomechanisch entnommen.



Before the deliberations half of the jurors were told:

Before you being your group discussion, think of yourself as a one man jury deciding the verdict in the case you just heard. In your opinion, is the defendant

Guilty?
Not Guilty?

The other half were told

Before you being your group discussion, think of yourself as a one man jury deciding the verdict in the case you just heard. In your opinion, what is the likelihood or probability that the defendant committed the criminal act for which he is charged?

For example: If you think that there is a (five out of ten chance or probability) that the defendant committed the act you will place a mark on the scale drawn below as follows:

10 9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0
out out
of 10 of 10

ON THE SCALE DRAWN BELOW MARK THE SPOT THAT
BEST DESCRIBES THE PROBABILITY YOU HAVE IN MIND:

I believe that there is a ——— out of ten chance that the defendant committed the criminal act for which he is charged:

10 9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0
out out
of 10 of 10

Figure 1

The first group of jurors was asked:

How does your verdict translate into a probability statement? Does it mean that there is a one in ten chance that the defendant committed the criminal act for which he is charged, or a six in ten chance, or a nine in ten chance, or a 9.5 in ten chance?

ONE THE SCALE DRAWN BELOW MARK THE SPOT THAT
BEST DESCRIBES THE PROBABILITY YOU HAVE IN MIND:

I believe that there is a ——— out of ten chance that the defendant committed the criminal act for which he is charged:

10 9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0
out out
of 10 of 10

The other half of the jurors, who had made the probability estimate first, were then told: Suppose you were asked on the basis of your probability statement to convict or acquit the defendant: would you find the defendant guilty or not guilty?

Figure 2

(Simon&Mahan, S.322)

(Simon&Mahan, S.321)

TABLE 1
QUALITATIVE AND QUANTITATIVE PREDELIBERATION
VERDICTS BY JURORS AND STUDENTS

Verdict	Students		Jurors	
	Qualitative	Quantitative	Qualitative	Quantitative
Guilty	"Qualitative form first"		"Qualitative form first"	
	70%	8.3	51%	7.8
Not Guilty	80%		48%	
	In = 44)		In = 37)	
Guilty	9%		35%	
	In = 44)		In = 37)	
Not Guilty	91%		62%	
	In = 44)		In = 32)	

TABLE 2
GROUP DECISIONS IN QUALITATIVE AND
QUANTITATIVE TERMS

Verdict	Students		Jurors	
	Qualitative	Quantitative	Qualitative	Quantitative
Guilty	"Qualitative form first"		"Qualitative form first"	
	5	4.0	1	6.5
Not Guilty	2		3	
	7.0		4.8	
Guilty	"Probability estimate first"		"Probability estimate first"	
	6	4.1	4	3.8
Not Guilty	1		1	
	4.1		3.8	

TABLE 3
QUALITATIVE AND QUANTITATIVE POSTDELIBERATION
VERDICTS BY JURORS AND STUDENTS

Verdict	Students		Jurors	
	Qualitative	Quantitative	Qualitative	Quantitative
Guilty	"Qualitative form first"		"Qualitative form first"	
	16%	8.3	32%	8.3
Not Guilty	83%		68%	
	4.8		4.8	
Guilty	"Probability estimate first"		"Probability estimate first"	
	5%	8.0	19%	8.3
Not Guilty	95%		81%	
	4.6		4.2	

(Simon&Mahan, S.323)

The first series of items asked:

1. What would the likelihood or probability have to be that a defendant committed the act for you to decide that he is guilty?
I would have to believe that it was a _____ out of ten chances that the defendant committed the act.
2. When a jury finds a defendant guilty after having been instructed that the burden of proof must be beyond a reasonable doubt, where was the probability, in the jury's mind that the defendant committed the act?
On the scale drawn (see sample of scale in Figure 2), mark the spot that best describes the probability you think the jury had in mind.
3. In a bench trial when a judge finds a defendant guilty beyond a reasonable doubt, where is the probability in the judge's mind that the defendant committed the act?
On the scale drawn (see sample of scale in Figure 2), mark the spot that best describes the probability you think the judge had in mind.

Figure 3

TABLE 4
PROBABILITY OF GUILT NECESSARY FOR DECIDING THAT
DEFENDANT IS GUILTY—REPORTED FOR SELF

Judges				Jurors				Students			
Prob- ability	Fre- quency	Cumula- tive Fre- quency	Prob- ability	Prob- ability	Fre- quency	Cumula- tive Fre- quency	Prob- ability	Prob- ability	Fre- quency	Cumula- tive Fre- quency	Prob- ability
0.0-0.0	01	-	0.0-0.0	06	-	16%	0.0-0.0	03	-	4%	0.0-0.0
5.0	02	1%	5.0	06	16%	5.0	02	01	-	4%	5.0
5.5	02	3%	5.5	02	18%	5.5	01	-	-	7%	5.5
6.0	08	4%	6.0	04	26%	6.0	01	-	-	7%	6.0
6.5	02	5%	6.5	01	26%	6.5	05	-	-	9%	6.5
7.0	14	8%	7.0	02	28%	7.0	01	01	14%	9%	7.0
7.5	23	14%	7.5	02	32%	7.5	01	01	14%	9%	7.5
8.0	58	31%	8.0	08	43%	8.0	09	30%	30%	30%	8.0
8.5	21	37%	8.5	02	46%	8.5	02	35%	35%	35%	8.5
9.0	68	57%	9.0	09	59%	9.0	21	55%	55%	55%	9.0
9.5	44	69%	9.5	03	64%	9.5	17	81%	81%	81%	9.5
10.0	106	100%	10.0	28	100%	10.0	30	100%	100%	100%	10.0
N.A.	04	-	N.A.	-	-	-	-	-	-	-	N.A.
Mode	10.0	-	Mode	10.0	-	-	Mode	10.0	-	-	Mode
Median	8.8	-	Median	8.8	-	-	Median	9.1	-	-	Median
Mean	8.9	-	Mean	7.9	-	-	Mean	8.9	-	-	Mean

(Simon&Mahan, S.324)

(Simon&Mahan, S.324)

TABLE 5
PUTATIVE PROBABILITY OF A DEFENDANT'S GUILT
"IN A JURY'S MIND"

Judges				Jurors				Students			
Prob- ability	Fre- quency	Cumula- tive Fre- quency	Prob- ability	Prob- ability	Fre- quency	Cumula- tive Fre- quency	Prob- ability	Prob- ability	Fre- quency	Cumula- tive Fre- quency	Prob- ability
0.0-0.0	02	-	0.0-0.0	04	-	16%	0.0-0.0	02	-	11%	0.0-0.0
5.0	06	2%	5.0	09	5.5	01	5.5	08	-	11%	5.0
5.5	03	3%	5.5	01	5.5	01	5.5	01	21%	32%	5.5
6.0	12	7%	6.0	03	6.0	03	26%	6.0	01	12%	6.0
6.5	01	8%	6.5	02	26%	6.5	06	-	-	-	6.5
7.0	18	12%	7.0	02	28%	7.0	02	31%	7.0	12	26%
7.5	30	27%	7.5	08	40%	7.5	08	40%	7.5	08	33%
8.0	57	40%	8.0	09	64%	8.0	14	49%	8.0	14	49%
8.5	18	49%	8.5	03	59%	8.5	10	60%	8.5	10	60%
9.0	63	64%	9.0	11	75%	9.0	19	82%	9.0	19	82%
9.5	20	74%	9.5	03	79%	9.5	07	90%	9.5	07	90%
10.0	86	100%	10.0	14	100%	10.0	09	100%	10.0	09	100%
N.A.	24	-	N.A.	02	-	-	N.A.	-	-	-	N.A.
Mode	10.0	-	Mode	10.0	-	-	Mode	9.0	-	-	Mode
Median	8.6	-	Median	7.8	-	-	Median	8.1	-	-	Median
Mean	8.7	-	Mean	7.7	-	-	Mean	8.0	-	-	Mean

TABLE 6
PUTATIVE PROBABILITY OF A DEFENDANT'S GUILT
"IN A JUDGE'S MIND"

Judges				Jurors				Students			
Prob- ability	Fre- quency	Cumula- tive Fre- quency	Prob- ability	Prob- ability	Fre- quency	Cumula- tive Fre- quency	Prob- ability	Prob- ability	Fre- quency	Cumula- tive Fre- quency	Prob- ability
0.0-0.0	01	-	0.0-0.0	05	-	14%	0.0-0.0	01	-	4%	0.0-0.0
5.0	02	1%	5.0	04	5.0	02	5.0	02	-	4%	5.0
5.5	02	2%	5.5	04	5.5	02	5.5	02	-	4%	5.5
6.0	06	3%	6.0	04	6.0	02	6.0	02	-	4%	6.0
6.5	03	4%	6.5	03	6.5	03	6.5	03	-	4%	6.5
7.0	09	7%	7.0	02	23%	7.0	02	23%	7.0	02	23%
7.5	20	12%	7.5	01	24%	7.5	04	14%	14%	14%	7.5
8.0	55	26%	8.0	05	32%	8.0	14	30%	30%	30%	8.0
8.5	23	35%	8.5	03	36%	8.5	05	35%	35%	35%	8.5
9.0	66	54%	9.0	12	64%	9.0	21	89%	89%	89%	9.0
9.5	43	67%	9.5	04	61%	9.5	19	81%	81%	81%	9.5
10.0	114	100%	10.0	26	100%	10.0	17	100%	10.0	17	100%
N.A.	07	-	N.A.	03	-	-	N.A.	-	-	-	N.A.
Mode	10.0	-	Mode	10.0	-	-	Mode	8.8	-	-	Mode
Median	8.9	-	Median	8.9	-	-	Median	8.8	-	-	Median
Mean	8.9	-	Mean	8.3	-	-	Mean	8.3	-	-	Mean

(Simon&Mahan, S.326)

TABLE 7
PROBABILITY OF CONVICTION BY PREPONDERANCE OF
EVIDENCE STANDARD, REPORTED FOR SELF

Judges			Jurors			Students		
Prob- ability	Pre- frequency	Comma- the two Pre- frequency	Prob- ability	Pre- frequency	Comma- the two Pre- frequency	Prob- ability	Pre- frequency	Comma- the two Pre- frequency
0.0-0.0	02	-	0.0-0.0	02	-	0.0-0.0	01	-
5.0	03	1%	5.0	03	1%	5.0	04	6%
5.5	184	56%	5.5	05	14%	5.5	01	7%
6.0	69	77%	6.0	02	19%	6.0	09	17%
6.5	14	81%	6.5	02	22%	6.5	03	21%
7.0	15	86%	7.0	04	28%	7.0	17	40%
7.5	17	91%	7.5	15	49%	7.5	11	53%
8.0	04	93%	8.0	12	67%	8.0	13	68%
8.5	04	94%	8.5	01	68%	8.5	04	72%
9.0	06	96%	9.0	12	88%	9.0	14	86%
9.5	04	97%	9.5	02	89%	9.5	05	94%
10.0	10	100%	10.0	08	100%	10.0	05	100%
N.A.	17		N.A.	-		N.A.	01	
Mode	5.5		Mode	7.5		Mode	7.0	
Median	5.4		Median	7.5		Median	7.4	
Mean	6.1		Mean	7.7		Mean	7.6	

TABLE 8
PROBABILITY OF CONVICTION BY PREPONDERANCE OF
EVIDENCE STANDARD, ASCRIBED TO "A JURY"

Judges			Jurors			Students		
Prob- ability	Pre- frequency	Comma- the two Pre- frequency	Prob- ability	Pre- frequency	Comma- the two Pre- frequency	Prob- ability	Pre- frequency	Comma- the two Pre- frequency
0.0-0.0	06	-	0.0-0.0	02	-	0.0-0.0	02	-
5.0	07	4%	5.0	05	10%	5.0	07	10%
5.5	146	50%	5.5	09	24%	5.5	02	12%
6.0	54	68%	6.0	02	27%	6.0	09	23%
6.5	17	73%	6.5	05	34%	6.5	06	30%
7.0	19	79%	7.0	09	48%	7.0	15	46%
7.5	31	89%	7.5	06	57%	7.5	09	57%
8.0	08	92%	8.0	11	73%	8.0	18	77%
8.5	09	94%	8.5	01	75%	8.5	06	84%
9.0	06	96%	9.0	10	90%	9.0	10	95%
9.5	02	97%	9.5	01	91%	9.5	01	96%
10.0	09	100%	10.0	06	100%	10.0	03	100%
N.A.	39		N.A.	02		N.A.	-	
Mode	5.5		Mode	8.0		Mode	8.0	
Median	5.5		Median	7.1		Median	7.2	

(Simon&Mahan, S.327)

TABLE 9
PROBABILITY OF CONVICTION BY PREPONDERANCE OF
EVIDENCE STANDARD, ASCRIBED TO "A JUDGE"

Judges			Jurors			Students		
Prob- ability	Pre- frequency	Comma- the two Pre- frequency	Prob- ability	Pre- frequency	Comma- the two Pre- frequency	Prob- ability	Pre- frequency	Comma- the two Pre- frequency
0.0-0.0	03	-	0.0-0.0	05	-	0.0-0.0	01	-
5.0	169	52%	5.0	04	14%	5.0	06	9%
5.5	70	72%	5.5	07	17%	5.5	01	9%
6.0	13	77%	6.0	07	18%	6.0	06	19%
6.5	13	81%	6.5	10	21%	6.5	02	18%
7.0	13	81%	7.0	02	21%	7.0	13	32%
7.5	22	87%	7.5	10	36%	7.5	10	46%
8.0	11	91%	8.0	05	59%	8.0	22	70%
8.5	03	92%	8.5	05	59%	8.5	05	70%
9.0	08	94%	9.0	08	72%	9.0	12	90%
9.5	08	96%	9.5	06	82%	9.5	05	95%
10.0	20	100%	10.0	12	100%	10.0	04	100%
N.A.	70		N.A.	03		N.A.	01	
Mode	5.5		Mode	10.0		Mode	8.0	
Median	5.5		Median	7.9		Median	7.6	
Mean	6.3		Mean	7.9		Mean	7.3	

(Simon&Mahan, S.328)

Der gleiche Fragebogen wurde für die individuellen Entscheidungen nach der Beratung und für die Gruppenentscheidungen ausgefüllt, wobei man die Teilnehmer wiederum in zwei Gruppen unterteilte, so daß ein Teil zuerst qualitative, der andere zuerst quantitative Urteile fällen mußte und dann umgekehrt. (Die Antworten sind in den Tabellen 1 bis 3 angegeben¹⁾).

Wie aus Tabelle 1 zu ersehen ist, waren Studenten eher geneigt, zugunsten des Angeklagten zu entscheiden als die Geschworenen. Beide, Studenten und Geschworenen, zeigten aber die gleiche Reaktion, wenn sie zuerst die Wahrscheinlichkeitsbeurteilung machen mußten: In beiden Fällen wurde viel häufiger zugunsten des Angeklagten entschieden. Aus der Tabelle ersieht man, daß bei den zuerst qualitativen Entscheidungen unter den Studenten 20% für schuldig und 80% für unschuldig, bei den Geschworenen 51% für schuldig und 49% für unschuldig plädierten; bei den zuerst quantitativen Beurteilungen dagegen waren nur noch 9% unter den Studenten und 38% unter den Geschworenen für schuldig. Auch können wir sehen, daß die Durchschnittswahrscheinlichkeit für die Schuld des Angeklagten (bei denjenigen, die für "schuldig" plädierten) bei den Studenten höher lag als bei den Geschworenen: 8.3 zu 7.8 bei einer "zuerst qualitativen" Entscheidung. Die gleichen Ergebnisse, nämlich daß Studenten vorsichtiger waren bei der Verurteilung des Angeklagten und daß die Übersetzung des subjektiven Gefühls bzw. Glaubens in eine Wahrscheinlichkeitsbewertung anscheinend einen gewissen Grad an Vorsicht zur Folge hatte, die sich dann zugunsten des Angeklagten auswirkte²⁾ sehen wir auch in der Tabelle 3, die die individuellen Antworten nach der Beratung angibt. Was allerdings die Wahrscheinlichkeitsbewertung betrifft, schien die Beratung einen nivellierenden Charakter zu haben (vgl. 8.3 bei Studenten bzw. 8.3 bei Geschworenen im "zuerst qualitativen" Fall und 8.0 bzw. 8.3 im "zuerst quantitativen" Fall).

1) Simon und Mahan haben die Wahrscheinlichkeitsskala von 0 bis 10 gewählt anstelle der gewöhnlichen von 0 bis 1. Anscheinend waren sie der Meinung, daß die Teilnehmer besser mit der ersten (etwas gröberen) als mit der zweiten umgehen können würden. Wie sie aber selbst schreiben (S.330), macht die Wahl einer Skala in der Wahrscheinlichkeitsverteilung und den Durchschnittswerten keinen wesentlichen Unterschied. Deswegen können wir ihre Werte jederzeit normieren, um eine mit der Wahrscheinlichkeitstheorie konsistente Beurteilung zu erhalten.

2) Vgl. Simon und Mahan, S.322. Gerade diese Tatsache scheint der Auffassung Tribes (vgl. S.129 dieser Arbeit) zu widersprechen, daß die Einführung einer Wahrscheinlichkeitsbeurteilung zur "Dehumanisierung" und dergl. führen würde.

Wir wollen die Ergebnisse aus den Tabellen 1 und 3 zusammenfassen:

1. Studenten sind in ihrer Beurteilung des Rechtsfalls vorsichtiger als Menschen, die repräsentativ aus allen Bevölkerungsschichten ausgesucht werden.
2. Die Übersetzung der subjektiven Gefühle in Wahrscheinlichkeitswerte führt dazu, sich mit dem Fall vorsichtiger und sogar 'humaner' zu befassen (der Angeklagte wird weniger häufig für schuldig befunden als im qualitativen Fall).
3. Wahrscheinlichkeitswerte, die für die Verurteilung ausreichen, sind im Durchschnitt viel geringer als man gewöhnlich unter einer "an Sicherheit grenzenden Wahrscheinlichkeit" versteht. (Bei den Geschworenen lagen sie im Durchschnitt im Falle der "zuerst quantitativen" Entscheidung sogar bei 6.8 (vgl. Tabelle 1)).
4. Der geringere Wahrscheinlichkeitswert (im Durchschnitt) für die Täterschaft (als Beweiswert für die Verurteilung) hat auf die Häufigkeit der Verurteilung keinen negativen Einfluß. Aus Tabelle 1 ist zu ersehen, daß im "zuerst quantitativen" Fall zwar die Durchschnittswahrscheinlichkeit für die "Schuld" (bei denjenigen, die für die Verurteilung plädieren) sowohl bei den Studenten wie auch bei den Geschworenen sinkt (jeweils 7.5 bzw. 6.8 gegenüber 8.3 bzw. 7.8 im "zuerst qualitativen" Fall), die Häufigkeit der Verurteilung deswegen aber nicht größer wird (diese sinkt sogar).

Aus Tabelle 2 (Gruppenentscheidungen) ersehen wie die gleichen Ergebnisse:

1. Studentengerichte neigen weniger dazu, den Angeklagten zu verurteilen;
2. Die "zuerst quantitative" Beurteilung führt zu einer günstigeren bzw. vorsichtigeren Einstellung gegenüber dem Angeklagten.

Interessant ist, daß der Wahrscheinlichkeitswert für die Schuld des Angeklagten im Falle der Aussetzung der Entscheidung bei den Geschworenen in den "zuerst qualitativen" Entscheidungen nur bei 4.8 lag (einem Wert, der eigentlich zum Freispruch ausreichen dürfte); bei den Studenten betrug dieser Wert 7.0 (bei diesem Wert wußte man also nicht, ob der Angeklagte verurteilt oder freigesprochen werden sollte).

Aus den Gruppenentscheidungen ersehen wir übrigens, daß nicht nur die Wahrscheinlichkeitsbeurteilung, sondern offensichtlich auch

die Beratung innerhalb einer Gruppe zu einer vorsichtigeren und 'humaneren' Einstellung gegenüber dem Angeklagten führt. (Das gleiche Ergebnis zeigt sich bei den individuellen Entscheidungen nach der Beratung (Tabelle 3)).

Es sind somit, wie wir sehen, drei Aspekte, die die Gerichtspraxis verbessern können: Die Beratung innerhalb einer Gruppe (die die individuellen Entscheidungen 'korrigiert'), die Gruppenentscheidung (d.h. wenn mehrere Personen über den Fall entscheiden und dann ihr gemeinsames Urteil präsentieren) und die Übersetzung der subjektiven Gefühle bzw. des Glaubens bzgl. des Falls in Wahrscheinlichkeitswerte.

Der zweite Teil der Untersuchung bestand darin, die Antworten der Studenten und der Geschworenen mit denen der Richter¹⁾ zu vergleichen. Alle Teilnehmer mußten wiederum einen Fragebogen ausfüllen (vgl. Fig. 3), in dem folgende Fragen standen:

1. Wie hoch muß für Sie die Wahrscheinlichkeit sein, daß der Angeklagte die Straftat begangen hat, damit Sie auf "schuldig" erkennen?
2. Wenn das Geschworenengericht den Angeklagten für schuldig befunden hatte, nachdem es instruiert worden war, daß das Beweiskriterium die "an Sicherheit grenzende Wahrscheinlichkeit" sein sollte, wie hoch war dann die Wahrscheinlichkeit für das Geschworenengericht, daß der Angeklagte die Tat begangen hatte?
3. Die gleiche Frage wie 2, diesmal allerdings in bezug auf die Richter.

(Die Antworten findet der Leser in den Tabellen 4, 5 und 6).

Die Ergebnisse zeigen, daß die zur Verurteilung des Angeklagten ausreichenden Wahrscheinlichkeiten im Intervall von 5.0 bis 10.0 liegen. Für manche Versuchspersonen würde sogar der Wert von 5.0 für die Verurteilung ausreichen (vgl. Tabelle 4: 1 Person unter den Richtern, 6 Personen unter den Geschworenen und 2 unter den Studenten); bei 6.0 liegen 3% aller Antworten der Richter, 25%(!) derer der Geschworenen und 7% derer der Studenten; bei 8.0 oder weniger liegen insgesamt fast 1/3 aller Antworten (vgl. 31% bei Richtern, 43% bei Geschworenen und 30% bei Studenten).

1) Der Aufsatz von Simon, der über eine ähnliche Untersuchung an Richtern berichtet, war mir leider während der Abfassung dieser Arbeit nicht zugänglich (R.J.Simon, Judges' translations of burdens of proof into statements of probability, in: Trial Lawyers' Guide).

Anschließend mußten die Teilnehmer das Beweiskriterium "mit überwiegender Wahrscheinlichkeit" in ein Wahrscheinlichkeitsurteil übersetzen. Es ist interessant festzustellen, daß die Antworten der Geschworenen und der Studenten ziemlich von denen der Richter abweichen (vgl. Tabellen 7, 8 und 9). Für die Richter betrug diese Wahrscheinlichkeit im Durchschnitt 5.5; bei den Studenten und Geschworenen lag sie dagegen etwa bei 7.5; d.h. für die letzten beiden Gruppen war der Unterschied zwischen den Beweiskriterien "mit an Sicherheit grenzender Wahrscheinlichkeit" und "mit überwiegender Wahrscheinlichkeit" viel geringer als für die Richter.

Zusammenfassung

Es sind v.a. die beiden folgenden in der Untersuchung von Simon und Mahan gewonnenen Ergebnisse, die für uns von Interesse sein können:

1. Die Wahrscheinlichkeitsbewertung des Sachverhalts führt zu einer vorsichtigeren bzw. 'milderen' Beurteilung (zugunsten des Angeklagten) als die qualitative (traditionelle).
2. Der individuelle Wahrscheinlichkeitswert für die Schuld, die für die Verurteilung des Angeklagten ausreichen würde, und der üblicherweise mit dem Ausdruck "mit an Sicherheit grenzender Wahrscheinlichkeit" umschrieben wird, ist weder nicht quantifizierbar noch liegt er, wie man manchmal in der Fachliteratur annimmt, bei 0.99...¹⁾, sondern variiert je nach der Beurteilungsperson (und je nach Fall) von 0.5 bis 1²⁾

Aber nicht nur die Wahrscheinlichkeitswerte für den Ausdruck "mit an Sicherheit grenzender Wahrscheinlichkeit" variieren, sondern v.a. die individuellen Beurteilungen ein und desselben Falls (vgl. Tabelle 1, wo der durchschnittliche Wahrscheinlichkeitswert für die Täterschaft von 0.3 bei den Geschworenen bei einer "zuerst quantitativen" Entscheidung (in diesem Fall wurde der Angeklagte

1) Vgl. S. 58 Fn.1, dieser Arbeit.

2) Wir normieren die von Simon & Mahan benutzte Wahrscheinlichkeitskala, um sie mit der Wahrscheinlichkeitstheorie in Einklang zu bringen; vgl. S. 153 Fn.1 dieser Arbeit.

für nicht schuldig befunden) bis zu 0.83 bei den Studenten bei einer "zuerst qualitativen" Entscheidung (in diesem Fall wurde der Angeklagte für schuldig befunden) variiert)¹⁾ Gerade die Untersuchung von Simon und Mahan (in der ein und derselbe Fall von viel mehr Personen beurteilt wurde, als das in der Gerichtspraxis üblich ist) hat uns gezeigt, wie Bildung, Alter, Einstellung, Hintergrundinformationen, mit einem Wort, Erfahrung, offensichtlich die Beurteilung beeinflussen. In besonders krasser Form zeigt Tabelle 1, wie solche Faktoren offenbar die Entscheidung mitkonstituieren: denn während nur 20% aller Studenten für die Schuld sprachen und 80% dagegen, erkannten unter den Geschworenen mehr als die Hälfte (51%) auf schuldig.

Wir dürfen somit (anhand der Untersuchung) annehmen, daß die Wahrscheinlichkeitsbeurteilung nicht nur zu einer vorsichtigeren und 'milderer' Gerichtspraxis führt²⁾ sondern auch die Tatsache aufdeckt, daß bestimmte Faktoren, die mit dem Fall nichts zu tun haben, die Urteilsbildung beeinflussen³⁾

1) was eine Bestätigung der Annahme ist, daß der intuitive Begriff der Wahrscheinlichkeit, der den juristischen Beurteilungen zugrunde liegt, einzig und allein durch den subjektiven Wahrscheinlichkeitsbegriff zu präzisieren ist.

2) Trotz dieser Vorteile waren jedoch 90% aller Richter, 62% aller Geschworenen und 69% aller Studenten gegen die Anwendung der Wahrscheinlichkeitstheorie in der Rechtspraxis (vgl. die Befragung von Simon&Mahan, S.329,330,Fn.8). Hauptargument war, daß Wahrscheinlichkeitsurteile (obwohl sie nichts anderes darstellen als die numerische Präzisierung der bereits vorhandenen subjektiven Gefühle und Glauben - Anm. von mir) nicht alle Faktoren bei der Feststellung der Schuld berücksichtigen könnten.

3) Eggleston, Beyond Reasonable Doubt (1977); S.7, berichtet von ähnlichen Ergebnissen wie bei Simon&Mahan, die von einer anderen Wissenschaftlerin erzielt wurden, vgl.: "In an investigation conducted by Miss Bridget Walsh, a student of Professor John Cohen of the Department of Psychology of Manchester University, it was found that one third of the subjects fixed a level below 0.7, one third thought the level should be above 0.9, while the remaining third chose a level between those figures" (gemeint ist der Wahrscheinlichkeitswert für die Schuld, der ausreichen würde, um den Angeklagten zu verurteilen).

2.2. Marshall und Wise¹⁾

In ihrem Aufsatz untersuchen Marshall und Wise, wie die persönliche Einstellung der Individuen zur Todesstrafe (Vorurteile, Skrupel und dergl.) sich auf die Entscheidung auswirkt. Sie kommen zum Ergebnis, daß Individuen, die für die Todesstrafe sind, auch dazu neigen, den Angeklagten öfter zu verurteilen. Marshall und Wise benutzen in ihrer Untersuchung einen entscheidungstheoretischen Ansatz. Wie ihre Untersuchung zeigt, waren die untersuchten Personen in ihren Entscheidungen mit dem Bayesianistischen Entscheidungsmodell i.d.R. konsistent.

Modell

Für die Bestimmung der Wahrscheinlichkeit der Schuld benutzen Marshall&Wise die von Finkelstein&Fairley²⁾ vorgeschlagene Bayes-Formel, wobei sie konsequent den subjektiven Wahrscheinlichkeitsbegriff (sowohl bei der Bestimmung der Apriori-Wahrscheinlichkeit als auch bei der Beurteilung des Beweismittelgewichts) verwenden. Darüber hinaus schlagen sie zum ersten Mal ein konsistentes entscheidungstheoretisches Modell für die juristische Urteilsbildung vor³⁾ das nicht nur mit dem subjektiven Wahrscheinlichkeitsbegriff, sondern auch mit einem subjektiven Nützlichkeitsbegriff arbeitet⁴⁾

Die Entscheidungssituation sieht nach Marshall&Wise so aus⁵⁾

Entscheidung	Tatsächlicher Zustand des Angeklagten	
	Schuld	Unschuld
schuldig	'Treffer'	'falscher Alarm'
unschuldig	'Fehlschluß'	'korrekte Ablehnung'

(vgl. die englischen Bezeichnungen: 'hit' $\hat{=}$ 'Treffer', 'false alarm' $\hat{=}$ 'falscher Alarm', 'miss' $\hat{=}$ 'Fehlschluß' und 'correct rejection' $\hat{=}$ 'korrekte Ablehnung').

1) C.R.Marshall&J.A.Wise, Juror Decisions and the Determination of Guilt in Capital Punishment Cases: A Bayesian Perspective, in: Utility, Probability and Human Decision Making (1975), S.257-269.

2) Vgl. S.124f dieser Arbeit.

3) Das von Kaplan erstmals für die Rechtspraxis vorgeschlagene Entscheidungsmodell ist in dem Sinne nicht korrekt, als es zwar 2 Sachverhalte und 2 Entscheidungsmöglichkeiten annimmt, jedoch nur 2 (statt 4) Konsequenzen berücksichtigt. (Kaplan, Decision Theory and the Factfinding Process (1968); zur Darstellung und Kritik dieses Modells siehe den II.Teil dieser Arbeit).

4) Es handelt sich hier um ein sehr einfaches Modell, das weder die Tatsache berücksichtigt, daß mehr als zwei Sachverhalte (z.B. Mord, Totschlag, minder schwerer Fall des Totschlags etc.) und entsprechend mehr als zwei Entscheidungen vorliegen können, noch restriktiv genug ist (so kann man darin die Entscheidung 'schuldig' ableiten, obwohl die Wahrscheinlichkeit für 'Schuld' kleiner ist als die für 'Unschuld'). Zu einer axiomatisch aufgebauten rationalen juristischen Entscheidungstheorie, die u.a. diese Nachteile vermeidet, siehe den II.Teil dieser Arbeit.

5) Marshall&Wise, S.260.

Marshall und Wise ordnen den Konsequenzen folgende Nützlichkeiten zu (Abk. für die Nützlichkeiten der Konsequenzen 'Treffer', 'falscher Alarm', 'Fehlschluß' und 'korrekte Ablehnung' sind jeweils U_{hit} , $U_{f.a.}$, U_{mis} , $U_{c.r.}$):

$$U_{hit} = U_{c.r.} = 0; \quad U_{mis} = -1 \quad \text{und} \quad U_{f.a.} = -x.$$

$P(G/E)$ ist die von Marshall&Wise benutzte Abkürzung für die Wahrscheinlichkeit der Schuld des Angeklagten aufgrund der Daten; und entsprechend $P(NG/E)$ eine solche für die Wahrscheinlichkeit der Unschuld aufgrund der Daten.

Um die Entscheidung 'schuldig' in diesem Modell abzuleiten, muß gelten:

$$P(G/E)(U_{hit}) + P(NG/E)(U_{f.a.}) \geq P(NG/E)(U_{c.r.}) + P(G/E)(U_{mis})^1)$$

Daraus folgt:

$$P(G/E)(U_{hit} - U_{mis}) \geq P(NG/E)(U_{c.r.} - U_{f.a.})$$

und (mit der Annahme der Gleichheit):

$$\frac{P(G/E)}{P(NG/E)} = \frac{U_{c.r.} - U_{f.a.}}{U_{hit} - U_{mis}}$$

Mit den obigen Werten erhalten wir:

$$\frac{P(G/E)}{P(NG/E)} = \frac{0 - (-x)}{0 - (-1)} = x.$$

Marshall&Wise bezeichnen $\frac{P(G/E)}{P(NG/E)}$ als das kritische Verhältnis für die Verurteilung. Dieses ist nach der obigen Ableitung gleich der 'Disutility', die das Subjekt der Verurteilung eines Unschuldigen beimißt.

Im folgenden wollen wir das Experiment von Marshall&Wise kurz schildern.

Die Untersuchung wurde an 159 Soziologiestudenten der Ohio State University durchgeführt. Jedem Teilnehmer wurde die Schilderung von 4 Rechtsfällen präsentiert, wobei in zweien davon ein Schwarzer und in den restlichen beiden ein Weißer angeklagt war. Man versuchte folgendes zu ermitteln:

1) Der Leser möge sich über entscheidungstheoretische Begriffe und einen entscheidungstheoretischen Formalismus im II. Teil dieser Arbeit informieren.

1. Zu welcher Gruppe (Gruppe A - ohne Skrupel gegen die Todesstrafe, Gruppe B - nicht ganz ohne Skrupel und Gruppe C - absolut dagegen) jeder Teilnehmer gehört¹⁾;
2. die subjektiven Wahrscheinlichkeitswerte, die die Teilnehmer jedem Beweismittel zuordnen würden;
3. die subjektiven Apriori- und Aposteriori-Wahrscheinlichkeitswerte (die letzteren unter Berücksichtigung aller Beweismittel), die die Teilnehmer der Täterschaft des Angeklagten zuordnen würden (diese Wahrscheinlichkeit wurde den Teilnehmern als "Grad des Glaubens bzgl. der Schuld bzw. Unschuld des Angeklagten" erklärt).

Ferner mußten die Teilnehmer einige Fragen beantworten, woraus man ihre 'Disutilities' für die Verurteilung eines Unschuldigen ersehen konnte. (Die Antwort auf diese Fragen ist identisch mit dem kritischen Verhältnis $(P(G/E)/P(NG/E))$, welches oben abgeleitet wurde). Die Aposteriori-Wahrscheinlichkeit, über der das Individuum entsprechend dem entscheidungstheoretischen Modell den Angeklagten verurteilen mußte, wurde als $x/(1+x)$ festgelegt (x ist die Antwort des Individuums zur Nützlichkeitsfrage).

Die Ergebnisse Marshall&Wise's²⁾

1. Auf die Häufigkeit der Verurteilung wirkte sich der Grad der Skrupellosigkeit aus (Personen ohne jegliche Skrupel bzgl. der Todesstrafe verurteilten häufiger als diejenigen mit Vorbehalten bzgl. einer solchen Strafe).
2. Die Auffassung, daß die Testpersonen Angeklagte aus Minderheitsgruppen (in der Untersuchung waren es Schwarze) wesentlich

1) Die Beantwortung dieser Frage ist insofern wichtig, als es in Amerika üblich ist, die Geschworenen, bevor man sie zur Beurteilung eines Kriminalfalls zuläßt, über ihre Einstellung zur Todesstrafe zu befragen. Diejenigen, die gegen die Todesstrafe sind, werden als "zu skrupulös" ausgeschlossen.

2) Da Marshall&Wise es versäumt haben, die in ihren Tabellen (Tabellen I bis V, Marshall&Wise, S.264ff) benutzten Bezeichnungen zu erklären und auch, wie sie zu den dort angegebenen Werten gekommen sind, ist es leider unmöglich, ihre Ergebnisse in befriedigender Weise auszulegen. Wir beschränken uns daher hier auf die Wiedergabe ihrer eigenen Interpretation dieser Ergebnisse (vgl. *ibid.*, S.264ff).

öfter verurteilen, wurde durch die Untersuchung nicht bestätigt¹⁾.
 3. Dagegen hat der Grad der Skrupellosigkeit der Testpersonen auf die Schätzung der Aposteriori-Wahrscheinlichkeit für die Schuld des Angeklagten einen Einfluß: Skrupellose Teilnehmer (Gruppe A) ordneten höhere Aposteriori-Wahrscheinlichkeiten für die Schuld zu als skrupelhafte (Gruppe B und C).

4. Die Auffassung, daß skrupellose Personen häufiger verurteilen, weil sie den Daten größeres Gewicht beimessen als skrupelhafte Personen, wurde durch die Untersuchung nicht bestätigt.

5. Hingegen hat der Grad der Skrupellosigkeit für die Zuordnung der Apriori-Wahrscheinlichkeiten eine Bedeutung²⁾. So benutzten skrupellose Personen höhere Apriori-Wahrscheinlichkeiten, daß der Angeklagte schuldig ist, als nicht skrupellose (das Verhältnis der Gruppen A:B, $p = 0.10$; A:C, $p = 0.055$).

6. Die subjektiven kritischen Verhältnisse $P(G/E)/P(NG/E)$ ($= x$) waren zu 80.5% konsistent mit den Urteilen (nur 13 Personen waren unfähig, in einer vernünftigen Weise ihre subjektiven Nützlichkeiten anzugeben).

Das Verhältnis der inkonsistenten Antworten zu allen Antworten hing mit der Skrupellosigkeit zusammen: So waren die skrupellosen Testpersonen häufiger inkonsistent in ihren Urteilen als die skrupelhaften.

Insgesamt überstieg die Anzahl der inkonsistenten Urteile auf 'schuldig' die der inkonsistenten Urteile auf 'unschuldig'. Dies gibt Anlaß zur Annahme, daß einige Personen, v.a. die skrupellosen, die Voraussetzung der Äquivalenz zwischen den Nützlichkeiten für beide Typen von korrekten Entscheidungen verletzten (bzw. daß einige Personen eine höhere Nützlichkeit in der Verurteilung eines Unschuldigen als im Freispruch eines Schuldigen sahen).

1) Man beobachtete sogar den umgekehrten Effekt, nämlich, daß die Testpersonen die weißen Angeklagten häufiger verurteilten als die schwarzen, was vielleicht darauf zurückzuführen ist, daß es sich bei den Testpersonen um Studenten handelte, also um eine Bevölkerungsgruppe, die gegenüber dem Rassismusvorwurf genug sensibilisiert ist.

2) Die Teilnehmer wurden gebeten, nur die Tatsache des Arrests und der Anklage bei ihrer subjektiven Schätzung der Apriori-Wahrscheinlichkeit für die Schuld des Angeklagten zu berücksichtigen, d.h. bevor überhaupt irgendwelche relevanten Beweisstücke vorgestellt wurden (!)

Die skrupellosen Testpersonen hatten außerdem ein viel kleineres kritisches Verhältnis bzgl. der Verurteilung als die skrupelhaften. (Dies ist eine unmittelbare Konsequenz der Zuordnung einer größeren Nützlichkeit zur Verurteilung eines Unschuldigen seitens dieser Person¹⁾).

Diskussion

Der Untersuchung von Marshall&Wise lag ein komplizierteres Modell zugrunde als derjenigen von Simon&Mahan: Während letztere die Testpersonen nur ihre Wahrscheinlichkeiten für die Schuld intuitiv schätzen ließen, mußten die Testpersonen in der Untersuchung von Marshall und Wise sowohl die Apriori-Wahrscheinlichkeiten und das Gewicht der Beweismittel als auch die Aposteriori-Wahrscheinlichkeit für die Schuld schätzen. Darüber hinaus mußten sie auch ihre subjektiven Nützlichkeiten für den Fall angeben, daß der Angeklagte irrtümlicherweise verurteilt würde, d.h. das Modell von Marshall&Wise betraf nicht nur die Wahrscheinlichkeitsschätzung der Sachverhaltshypothese, sondern darüber hinaus eine rationale Wahl der Entscheidung. Wie die Untersuchung von Marshall und Wise gezeigt hat, bereitete den meisten Teilnehmern der Umgang mit dem Modell keine besonderen Schwierigkeiten; 80.5% aller Testpersonen waren in ihren Antworten mit diesem Entscheidungsmodell konsistent.

Außerdem zeigte diese Untersuchung erneut auf²⁾, daß die persönliche Einstellung (Vorurteile, Skrupel und dergl.) eine nicht zu unterschätzende Rolle bei der Entscheidungsbildung spielt. So waren skrupellose Personen in der Untersuchung (man hat die persönliche Einstellung nur in bezug auf einen einzigen Faktor (die Todesstrafe) untersucht) viel eher zu negativen Urteilen geneigt: Sie verurteilten häufiger, ordneten höhere Apriori-Wahrscheinlichkeitswerte für die Schuld zu (und somit auch höhere Aposteriori-Wahrscheinlichkeitswerte aufgrund der Daten), hatten ein kleineres kritisches Verhältnis zu der Verurteilung und waren öfter in ihren Antworten inkonsistent als skrupelhafte Testpersonen. Da man in der amerikanischen Rechtspraxis gerade skrupellose Personen als Geschworene aussucht (d.h. die Personen, die nicht gegen die Todesstrafe sind), setzt vielleicht die Untersuchung von Marshall&Wise ein Zeichen, diese Konvention in amerikanischen Gerichten neu zu überdenken.

1) Vgl. Marshall&Wise, S.268: "...Some even go so far as to prefer the convicting of an innocent man to the acquittal of a guilty man!"

2) Vgl. Simon&Mahan, S.150ff dieser Arbeit.

2.3. Eggleston

Bereits in seinem 1963 erschienenen Aufsatz¹⁾ befaßt sich Eggleston mit den beiden Beweiskriterien "mit an Sicherheit grenzender Wahrscheinlichkeit" und "mit überwiegender Wahrscheinlichkeit". Gleich zu Anfang zitiert er den australischen Richter L.J.Denning²⁾:

"As Best C.J. and many other great judges have said, "in proportion as the crime is enormous, so ought the proof to be clear". So also in civil cases, the case may be proved by a preponderance of probability but there may be degrees of probability within that standard. The degree depends on the subject-matter".

Für uns ist jedoch folgende Äußerung Dennings viel wichtiger, die - obwohl sie 'nur' theoretisch ist - der Auffassung widerspricht, daß mit dem Kriterium "mit an Sicherheit grenzender Wahrscheinlichkeit" lediglich ein zu vernachlässigender Zweifel oder - will man diesen Zweifel doch quantifizieren - ein Wert nahe bei 1 gemeint ist³⁾. Wie Denning schreibt, kann jeder Wert im Intervall von 0.51 bis 1 diese Wahrscheinlichkeit sein⁴⁾. Mit Dennings Worten⁵⁾:

"So the phrase "reasonable doubt" takes the matter no further. It does not say that the degree of probability must be as high as 99 per cent, or as low as 51 per cent. The degree required must depend on the mind of the reasonable and just man who is considering the particular subject-matter. In some cases 51 per cent would be enough, but not in others. When this is realised, the phrase "reasonable doubt" can be used just as aptly in a civil case or a divorce case as in a criminal case..."

Auch in einem weiteren Aufsatz⁶⁾ bemüht sich Eggleston um die Beantwortung verschiedener Fragen, die im Zusammenhang mit den beiden Beweiskriterien auftauchen. Diese Fragen und ihre Klärung tragen zum größten Teil einen philosophischen Charakter. Wir wollen uns jedoch hier nicht mit seiner philosophischen Analyse auseinandersetzen, sondern lediglich auf den Teil des Aufsatzes eingehen, in dem er die Ergebnisse einer empirischen Untersuchung bzgl. des Vergleichs der drei Beweiskriterien "mit Sicherheit", "mit an Sicherheit grenzender Wahrscheinlichkeit" und "mit überwiegender Wahrscheinlichkeit" schildert⁷⁾.

1) Sir R.Eggleston, Probabilities and Proof, in: 4 Melbourn Univ. Law Rev. (1963), S.180-211.

2) Der Rechtsfall 'Bater v. Bater' (1951), S.35,37; Eggleston, S.180

3) Auf S. 58 haben wir zwischen zwei Explikationen des theoretischen Begriffs "mit an Sicherheit grenzender Wahrscheinlichkeit" unterschieden: Die erste Explikation käme der Kontextelimination des Wahrscheinlichkeitsbegriffs gleich und die zweite hätte die Angabe eines numerischen Wertes (wie z.B. 0.99...) zur Folge. In der Untersuchung, die Eggleston beschreibt, wird wiederum gezeigt (was die nachfolgende Behauptung Dennings bestätigt), daß der empirische Begriff "mit an Sicherheit grenzender Wahrscheinlichkeit" mit keinem von beiden identifiziert werden kann.

4) Auch die bereits geschilderte Untersuchung von Simon&Mahan bestätigt diese Auffassung.

5) Eggleston, S.180.

6) Sir R.Eggleston, "Beyond Reasonable Doubt", in: 4 Monash Univ. Law Rev. (1977), S.1-22.

7) Vgl. ibid, S.9ff.

Eggleston schildert ein Projekt, das durch die Wissenschaftler der London School of Economics durchgeführt wurde und dessen Ziel es war, die Auswirkung verschiedener Beweismittel und verschiedener Beweiskriterien auf die Urteilsbildung zu untersuchen. Auch diese Untersuchung (wie die von Simon & Mahan) zeigte, daß Laienrichter dazu tendieren, die Beweiskriterien "mit an Sicherheit grenzender Wahrscheinlichkeit" und "mit überwiegender Wahrscheinlichkeit" für weniger verschieden voneinander zu betrachten als man gewöhnlich annimmt. Wichtig ist jedoch das Ergebnis, daß die Tests einen bedeutenden Unterschied zwischen den Beweiskriterien "mit Sicherheit" und "mit an Sicherheit grenzender Wahrscheinlichkeit" aufdeckten¹⁾ vgl. die Tabelle unten²⁾

NUMBERS AND PERCENTAGES FOR GUILTY (G) AND NOT GUILTY (NG)

Cases	Beyond Reasonable Doubt		Sure and Certain		More Likely than Not	
	G	NG	G	NG	G	NG
Theft Number	43	94	32	58	42	50
% G	31		33		46	
Rape 1st Defendant Number	94	48	39	32	30	14
% G	66		33		68	
Rape 2nd Defendant Number	46	96	13	58	18	26
% G	32		18		41	
All Cases Number	183	238	84	148	90	90
% G	43		36		50	

¹ Includes "Guilty of Attempted Rape".

Schilderung des Experiments:

Einer Gruppe von Laienrichtern wurden zwei Fälle präsentiert: in einem ging es um einen Fleischdiebstahl, in dem anderen um eine Vergewaltigung, an der zwei Täter beteiligt waren (und dementsprechend zwei Angeklagte vor Gericht standen). Jede Testperson wurde aufgefordert, ihr Urteil einmal nach der Schilderung des Falls und das andere Mal nach Beratung mit den anderen Beteiligten anzugeben.

1) Die etablierte Meinung ist die, daß beide Beweiskriterien im Grunde genommen dasselbe besagen, 'nur', daß das zweite (das in der letzten Zeit in der Rechtsprechung als Urteilsformel immer mehr an Bedeutung gewinnt) explizit "erkenntnistheoretische Zweifel, die prinzipiell immer bestehen" im Wortlaut berücksichtigt.

Der "bedeutende Unterschied zwischen den Beweiskriterien "mit Sicherheit" und "mit an Sicherheit grenzender Wahrscheinlichkeit" ist auch aus einem anderen Grund interessant: Hier zeigt sich, daß in einer Situation unter Unsicherheit (Risiko) viele Personen in der Praxis sich in dem Sinn irrational verhalten, als sie in dieser eine solche unter Sicherheit sehen.

2) Die Tabelle ist dem obengenannten Aufsatz von Eggleston (S.9) photomechanisch entnommen.

ben¹⁾

Als interessante Nebenergebnisse dieser Untersuchung führt Eggleston folgende an²⁾

1. Einigen Geschworenengerichten teilte man mit, daß der Angeklagte bereits mehrmals verurteilt worden wäre, wobei man wiederum einen Teil der Gerichte bat, diese Tatsache nicht zu berücksichtigen, während man dem anderen Teil keine diesbezügliche Aufforderung machte. In einigen Fällen erfolgten die früheren Verurteilungen für ähnliche Tatbestände, in anderen für grundverschiedene von dem, der zur Entscheidung anstand. Forderte man die Testpersonen nicht auf, dies für den gegenwärtigen Fall unberücksichtigt zu lassen, so stieg im ersten Fall die Zahl der Verurteilungen, im letzteren dagegen fiel sie rapide, so als wehrten sich die Beteiligten gegen einen - ihrer Ansicht nach ungeeigneten - Versuch, sie durch die Bezugnahme auf frühere Verurteilungen zu beeinflussen.

2. In dem Vergewaltigungsfall führte die Mitteilung von früheren Verurteilungen des zweiten Angeklagten wegen ähnlicher Delikte zu einem Anstieg der Zahl der Verurteilungen des ersten Angeklagten, und zwar auch dann, wenn gegen diesen keine solchenehemaligen Verfahren angeführt waren. Andererseits verringerte sich die Zahl der Verurteilungen, wenn man die Testpersonen bat, diese früheren Verurteilungen außer acht zu lassen.

Kehren wir nun zu der obigen Tabelle zurück. Aus ihr ersehen wir, daß die Aufforderung an die Laienrichter, sie müßten bei der Verurteilung "sicher" sein, zu einer Verringerung der Zahl der Verurteilungen (im Vergleich zum Kriterium "mit an Sicherheit grenzender Wahrscheinlichkeit") führte, während die Aufforderung, den Fall "mit überwiegender Wahrscheinlichkeit" zu entscheiden, eine Erhöhung dieser Zahl bewirkte. Die Zunahme der Verurteilungen aufgrund des Beweiskriteriums "mit überwiegender Wahrscheinlichkeit" ist jedoch nicht so groß, wie man erwarten würde. Speziell in dem Fall, wo der erste Angeklagte wegen Vergewaltigung verurteilt wurde, ist kaum ein Unterschied zwischen den beiden Beweiskriterien "mit an Sicherheit grenzender Wahrscheinlichkeit" und "mit überwiegender Wahrscheinlichkeit" festzustellen.

1) Den Testpersonen wurde mitgeteilt, was sie unter den verschiedenen Beweiskriterien verstehen sollten, so unter "beyond reasonable doubt": "You should be sure beyond reasonable doubt and by a reasonable doubt I mean not a fanciful doubt, but such a doubt that might affect you in daily business or domestic decision" (im Diebesfall, ähnlich im Vergewaltigungsfall, vgl. Eggleston, S.11); unter "sure and certain": "Before you convict you must feel sure and certain on the evidence you have heard that the accused is guilty" (im Diebstahlfall, ähnlich im Vergewaltigungsfall, ibid, S.11). Leider führt Eggleston keinen Beleg ähnlicher Art für das Kriterium "mit überwiegender Wahrscheinlichkeit" an.

2) Vgl. Eggleston, S.9f.

Dieses Ergebnis ist, wie Eggleston bemerkt (S.10), jedoch mit Vorsicht zu genießen; Denn wenn sehr viele Beweismittel gegen den Angeklagten sprechen, werden auch die Endergebnisse aufgrund dieser beiden Beweiskriterien viel weniger verschieden ausfallen, als wenn nicht so viele und nicht so starke Beweismittel gegen den Angeklagten vorliegen; vgl.: "To take the extreme case, if the evidence is overwhelming, the rate of convicting would be 100% whichever form of instruction was given to the jury. Where the case is closer to the borderline, "more likely than not" might produce a substantial number of convictions, but "beyond reasonable doubt" hardly any".

Auch wenn Eggleston das Ergebnis durch ein plausibles Argument abzuschwächen versucht, bestätigen andere Untersuchungen¹⁾, daß der Unterschied zwischen den beiden Beweiskriterien in der Praxis viel geringer ausfällt, als man gewöhnlich annimmt.

Aus der Tabelle ersieht man also, daß anscheinend viele Testpersonen einen Unterschied zwischen den Kriterien "mit an Sicherheit grenzender Wahrscheinlichkeit" und "mit Sicherheit" machten (die erstere wird für geringer gehalten als die letztere), was der Auffassung widerspricht, daß das erste Beweiskriterium lediglich "(erkenntnistheoretische) Zweifel beinhaltet, die immer bestehen, wie sicher man auch ist", d.h. daß es prinzipiell keinen Unterschied zwischen diesen beiden Kriterien in der Praxis gibt.

Zusammenfassung

Die von Eggleston geschilderte empirische Untersuchung hat gezeigt, daß 1) der Unterschied zwischen den Beweiskriterien "mit an Sicherheit grenzender Wahrscheinlichkeit" und "mit Sicherheit" in der Praxis viel bedeutender ist als man in der Fachliteratur behauptet und anscheinend nicht allein durch "erkenntnistheoretische Zweifel, die immer bestehen" zu rechtfertigen ist; und 2) der Unterschied zwischen den Kriterien "mit an Sicherheit grenzender Wahrscheinlichkeit" und "mit überwiegender Wahrscheinlichkeit" insgesamt geringer ausfällt, als man gewöhnlich annimmt (wobei man bei diesem Ergebnis, wie Eggleston bemerkt, berücksichtigen muß, ob im aktuellen Fall viele und starke Beweismittel gegen den Angeklagten vorliegen oder nur wenige und schwache. Macht es in der ersten Situation praktisch keinen Unterschied, aufgrund welchen Kriteriums man den Fall beur-

1) So zeigt die Tabelle 4 (Simon&Mahan), daß der Wahrscheinlichkeitswert für die "an Sicherheit grenzende Wahrscheinlichkeit" im Intervall von 0.5 bis 1, bei 25% aller Geschworenen sogar nicht über 0.6 liegt. Der Durchschnitt für diesen Ausdruck liegt bei Geschworenen bei 0.86, ein solcher für die "überwiegende Wahrscheinlichkeit" bei 0.75 (vgl. Tabelle 7); ähnliche Ergebnisse teilt Eggleston selber mit (vgl. S.156, Fn.3 dieser Arbeit).

teilt, so kann man mit Recht vermuten, daß ein bestimmtes Kriterium aufgrund einer schwachen Beweisgrundlage die Entscheidung beeinflussen wird. Es ist daher nicht so sehr das Beweiskriterium als vielmehr die Beweisgrundlage, die auf die Entscheidung einwirkt. Die Diskussion um das Beweiskriterium ist somit irrelevant, wenn viele Beweismittel gegen den Angeklagten vorliegen, dagegen relevant, wenn es wenige bzw. diese unzureichend sind).

Auf eine weitere Untersuchung sei hier nur ganz kurz eingegangen.

1.4 Nagel&Neef¹⁾

Auf S.196 schreiben die beiden Autoren:

"In general, the threshold probabilities²⁾ tend to be quite low compared to the .90 figure that the law seems to require and the close-to-.90 figure that most of the respondents said they thought the law requires or ought to require. The actual threshold probabilities tend to average about .55 and thus were no different from what one might expect to find if the respondents were operating under the civil case standard of by a preponderance or a majority of the evidence, rather than the criminal case standard of beyond a reasonable doubt". (Hervorhebung im Text)

und weiter (S.199f):

"In summary, our preliminary causal findings tend to indicate the following. (1) In general, the threshold probabilities tend to be substantially lower than the .90 level implied by Blackstone and explicitly stated by judges in direct questioning, regardless of the type of jurors or cases which we surveyed. (2) In specific types of cases, the threshold probabilities tend to differ substantially from one group of potential jurors to another. (3) Jurors often may decide first on whether to convict or acquit and then an appropriate probability of guilt and a conviction standard. Those findings raise the questions of (1) how can jurors be better instructed to apply higher threshold probabilities, (2) how jurors can be better instructed in order to decrease the differences across different background and attitudinal groups, and (3) how jurors can be better instructed so as to increase the likelihood that they will consciously think about the defendant's probability of guilt and the conviction standard before, not after, deciding how to vote".

Die folgende Stelle aus Nagel&Neef (S.200f) illustriert exemplarisch, wie unterschiedliche Wahrscheinlichkeiten, die bei einer bestimmten Straftat für die Verurteilung ausreichen, angesetzt werden können, je nachdem, wer und in welcher Weise von dieser Straftat betroffen ist; außerdem wird hier gezeigt, daß eine geeignete Instruierung und Aufklärung (die evtl. auch die Gründe für das unterschiedliche Verhalten aufdeckt) die Unterschiede in den Wahrscheinlichkeiten zum Teil nivellieren kann. Im vorliegenden Fall geht es um eine Vergewaltigung; die Befragten sind Frauen und Männer. (Wie Nagel&Neef (S.197) berichten, haben ihre früheren Befragungen gezeigt, daß Männer bei Vergewaltigungsdelikten ein höhe-

1) S.S.Nagel&M.G.Neef, Decision Theory and the Legal Process (1979).

2) "threshold probability" P^* ist ein Beweismaß mit folgender Eigenschaft: ist die aktuelle Wahrscheinlichkeit P für die Schuld des Angeklagten höher als P^* , dann wird der Angeklagte verurteilt, ist sie kleiner als P^* , dann wird der Angeklagte freigesprochen; gilt $P^*=P$, dann sind Verurteilung und Freispruch gleichwertig. Nach Nagel&Neef soll in diesem Fall die Entscheidung durch einen Münzwurf erreicht werden (vgl. Nagel&Neef, S.189).

res Beweismaß (d.h. eine höhere Wahrscheinlichkeit für die Schuld des Angeklagten) und eine geringere Nützlichkeit für die Verurteilung eines Unschuldigen ansetzten als dies bei Frauen der Fall war. Ein interessantes Nebenergebnis ist hier zu erwähnen: Je liberaler eine Frau war, desto geringer war auch ihr Beweismaß und ihre Sensibilität, d.h. desto höher die Nützlichkeit für die Verurteilung eines Unschuldigen in solchen Fällen).

Nach der Instruierung¹⁾ war folgendes zu beobachten (S.200f):

"Although the above equalizing type of instruction does seem to reduce divergence, as indicated by our rough pilot study, that type of instruction does not substantially increase conformity with the legal notion that the threshold probability should be about .90. More specifically, what seems to be happening is that males in rape cases who are uninstructed start out with an average threshold probability of about .70, and females who are uninstructed start out with an average threshold probability of about .50. If one then takes another random group of males and gives them the equalizing instruction, their threshold probability tends to drop to about .60; and if one takes another random group of females and gives them the equalizing instruction, their threshold probability tends to rise to about .60. Both groups, however, are substantially below the .90 figure".

Bei einem weiteren Experiment wurden die befragten Studenten in vier Gruppen unterteilt: (1) in diejenigen, die keine Instruktionen bekamen, (2) in diejenigen, die instruiert waren, als Beweismaß für die Verurteilung die "an Sicherheit grenzende Wahrscheinlichkeit" zu benutzen, (3) in diejenigen, die instruiert wurden, sich bei einer Verurteilung am Beweismaß einer Wahrscheinlichkeit von 0.9 zu richten, und (4) in diejenigen, denen gesagt wurde, daß sie nach dem Kriterium "es ist besser, 10 Schuldige freizusprechen als einen Unschuldigen zu verurteilen" entscheiden sollten. Die Ergebnisse dieser Befragung seien hier in einer Tabelle angeführt:

	Männer	Frauen
keine Instruktion	0.65	0.45
"mit an Sicherheit grenzende W."	0.75	0.60
"0.9 Wahrscheinlichkeit"	0.85	0.75
"10 gegen 1"	0.90	0.90

(in den Spalten sind die durchschnittlichen Wahrscheinlichkeitswerte für die Schuld angegeben, die (nach dem jeweiligen Beweismaß) ausreichen sollten, den Angeklagten wegen der Vergewaltigung zu verurteilen):

Nagel&Neef resümieren (S.201): "That pattern is an oversimplification, but it illustrates the idea that threshold-raising instructions can have the simultaneous effect of raising the probability of threshold and equalizing divergent probability thresholds. This pattern also points up that one is likely to find that merely using words like beyond a reasonable doubt does not sufficiently raise the threshold probability..." (Hervorhebung im Text)²⁾

1) Die folgendermaßen lautete (Nagel&Neef, S.200): "Social science research tends to show that males are more reluctant to convict rape defendants than females. When you go into the jury room if you are a male and you feel reluctant to convict, ask yourself whether your reluctance is based on the facts of the case or on your identifying with the defendant. If you are a female and you feel relatively willing to convict, ask yourself whether your willingness is based on the facts of the case or on your identifying with the victim".

2) Auch andere von Nagel&Neef angeführte Beispiele (S.197ff) zeigen, wie soziale bzw. berufliche Zugehörigkeit, Geschlecht, Intelligenz, Ausbildung, Einstellung (liberal/konservativ) die "threshold"-Wahrscheinlichkeit und die Wünschbarkeit der Verurteilung von Unschuldigen bei verschiedenen Verbrechenarten beeinflussen.

Zusammenfassung der Ergebnisse

Wir fassen die wichtigsten Ergebnisse aus den vier in diesem Abschnitt besprochenen Untersuchungen noch einmal zusammen, wobei wir diesen Ergebnissen einige übliche Einwände gegen die Anwendung der Wahrscheinlichkeitstheorie auf die Rechtspraxis gegenüberstellen:

1. Laien und Richter sind imstande, auch wenn sie es ungern tun (vgl. S. 156, Fn. 2), ihre subjektiven Urteile zu quantifizieren (vgl. Simon&Mahan).

Somit ist der Einwand gegen die Anwendung der Wahrscheinlichkeitstheorie, daß die Richter keine numerischen Wahrscheinlichkeitswerte ihren qualitativen Urteilen zuordnen könnten, nicht gerechtfertigt.

2. Die meisten Personen sind 'sogar'(!) in ihren subjektiven Urteilen mit dem Bayes-Ansatz konsistent (vgl. Marshall&Wise: die Rate der mit dem Bayes-Modell konsistenten Antworten bei allen Testpersonen betrug 80.5%, wobei der Bayes-Ansatz ein viel komplizierteres Wahrscheinlichkeitsmodell darstellt, als lediglich die intuitive Zuordnung von Wahrscheinlichkeitswerten zu Endurteilen (das Modell von Simon&Mahan). Denn die Individuen müssen hier nicht nur die Apriori- und Aposteriori-Wahrscheinlichkeitsbewertung, sondern auch eine solche der Daten vornehmen. Ein ähnliches Modell für die Ermittlung der "threshold-probability" schlagen Nagel&Neef vor).

Der Einwand gegen die Anwendung der Wahrscheinlichkeitstheorie, daß Juristen "anders denken" als es die Wahrscheinlichkeitstheorie verlangt (Vgl. Tribe, Cohen¹) wird also nicht bestätigt.

3. Das Denken in Wahrscheinlichkeiten führt zu einer vorsichtigeren Beurteilungspraxis (zugunsten des Angeklagten) als die herkömmliche in qualitativen "schuldig"/"unschuldig"-Urteilen (Simon&Mahan).

Somit ist der Einwand, daß die Wahrscheinlichkeitstheorie zur "Dehumanisierung", "Ungerechtigkeit" und dergl. (Tribe) führe, nicht gerechtfertigt.

4. Die Untersuchungen zeigen, daß unterschiedliche Menschen ein und denselben Fall unterschiedlich beurteilen (Simon&Mahan; 91% aller Studenten in "probability estimate first" sind für "unschuldig" während 51% aller Geschworenen in "qualitative form first" für "schuldig" plädieren (wobei es sich um ein und denselben Fall handelt) und der durchschnittliche Wahrscheinlichkeitswert für die Täterschaft liegt bei Geschworenen in "probability estimate first" bei 0.3²) während derjenige der Studenten für den gleichen Fall in "qualitative form first" 0.83³ beträgt (vgl. Tabelle 1); Marshall&Wise: die sog. skrupellosen Testpersonen ordneten sowohl höhere Apriori- wie auch höhere Aposteriori-Wahrscheinlichkeiten zu als die weniger skrupellosen in ein und demselben Fall; vgl. auch Nagel&Neef).

Somit ist u.a. bestätigt, daß der intuitive Wahrscheinlichkeitsbegriff der solchen Beurteilungen zugrunde liegt, nicht objektiv und nicht als statistischer Wahrscheinlichkeitsbegriff (im Sinn der relativen

1) auf Cohen wird im nächsten Paragraphen näher eingegangen.

2) im Fall des Freispruchs.

3) im Fall des Schuldspruchs.

Häufigkeit) explizierbar ist, sondern einzig und allein durch den Begriff der subjektiven Wahrscheinlichkeit präzisiert werden kann.

5. Der Begriff "mit an Sicherheit grenzender Wahrscheinlichkeit" hat anscheinend in der Praxis eine ganz andere Bedeutung als in der Fachliteratur. Keineswegs ist darunter ein objektives Kriterium zu verstehen, das im Prinzip das gleiche besagt wie die Redewendung "mit Sicherheit", wobei "nur erkenntnistheoretische Zweifel bestehen"¹⁾

Was darunter verstanden wird, variiert von Fall zu Fall und von Individuum zu Individuum (Simon&Mahan: bei 1/3 aller Geschworenen lag dieser Wert nicht über 0.75; vgl. auch Eggleston, Nagel&Neef). Es ist daher fraglich, ob man diesen Begriff im herkömmlichen Sinn (als "fast-Sicherheit", wobei "fast" gegen Null "strebt") benutzen soll, oder stattdessen als diejenige Wahrscheinlichkeit ansieht, die im konkreten Fall dem beurteilenden Richter ausreichen würde, den Angeklagten zu verurteilen²⁾

6. Der Unterschied zwischen den juristischen Beweiskriterien "mit an Sicherheit grenzender Wahrscheinlichkeit" und "mit überwiegender Wahrscheinlichkeit" fällt geringer aus, als man gewöhnlich meint (Simon&Mahan, Eggleston, Nagel&Neef). Man kann mit Recht annehmen, daß die beiden Beweiskriterien zu demselben Ergebnis führen, wenn die Informationsbasis (Beweisgrundlage) gegen den Angeklagten stark ist. In diesem Fall ist es daher irrelevant, welches Beweiskriterium man der Entscheidung zugrunde legt. Relevant wird diese Frage erst dann, wenn die Beweisgrundlage schwach ist.

7. Die Wahrscheinlichkeitsanalyse deckt explizit auf, daß Hintergrundinformationen, Wissen über den Angeklagten, die persönliche Einstellung (Vorurteile, Skrupel und dergl.), aber auch die Bildung, Intelligenz, Geschlecht des Individuums eine nicht zu unterschätzende Rolle bei der Urteilsbildung spielen (Marshall&Wise, Eggleston, Simon&Mahan, Nagel&Neef; die Variation der durchschnittlichen Wahrscheinlichkeitswerte von 0.83 bis 0.3 für die Täterschaft im gleichen Fall (Simon&Mahan, Tabelle 1) muß offenbar zum größten Teil auf diese individuellen Unterschiede zurückgeführt werden)).

Summiert man diese sieben Punkte, so ist der Wert der (subjektiven) Wahrscheinlichkeitstheorie sowohl bei der rationalen (konsequenten und konsistenten) Urteilsbildung wie auch bei der Aufdeckung von individuellen Unterschieden bei der Beurteilung juristischer Fälle offensichtlich (empirisch belegt).

1) Vgl. dazu auch eine höchst interessante Stelle bei Nagel&Neef; S.201f: "Few jurisdictions provide for instructing jurors as to the meaning of beyond a reasonable doubt. The justifications offered for not providing such instructions tend to emphasize that attempting to define a conviction standard will merely confuse or bias jurors. The real justification might be that legislators and state supreme court judges who promulgate judicial instructions may be aware that giving more explicit precision to the notion of "beyond a reasonable doubt" will tend to increase the probability of threshold and make convictions more difficult".

2) Es ist überflüssig zu betonen, daß ich die letztgenannte Interpretation vorschlage, die meiner Ansicht nach, und wie ich es in diesem Paragraphen herauszustellen versuchte, den Beurteilungsvorgang adäquater wiedergibt.

§3. Weitere Diskussion des Wahrscheinlichkeitsbegriffs: Cohen

Waren es ca. 15 Jahre lang ausschließlich Aufsätze¹⁾, in denen die juristische Diskussion um den Wahrscheinlichkeitsbegriff bzw. die Anwendung der Wahrscheinlichkeitstheorie auf die Rechtspraxis ausgetragen wurde, so erschienen 1977-78 nacheinander erstmals drei Bücher zu diesem Thema. Diese sind: "The Probable and the Provable" (1977) von L.J.Cohen, "Quantitative Methods in Law" (1978) von M.O.Finkelstein und "Evidence, Proof and Probability" (1978) von Sir R.Eggleston.

Während Finkelstein und Eggleston davon ausgehen, daß es sich bei dem Wahrscheinlichkeitsbegriff im Bereich des juristischen Rasonierens um den herkömmlichen Begriff der Wahrscheinlichkeit handelt, vertritt Cohen die Auffassung, daß der juristischen Intuition ein andersartig strukturierter Wahrscheinlichkeitsbegriff zugrunde liegt. In seinem Buch stellt er die Prinzipien dieses Wahrscheinlichkeitsbegriffs auf (die er im Unterschied zu den herkömmlichen Kolmogoroff-Axiomen - von ihm Pascalsche Prinzipien genannt - als Baconsche bezeichnet²⁾) und entwickelt eine, wie er mißverständlichlicherweise sie nennt, "induktive Wahrscheinlichkeitstheorie"³⁾, der diese Baconschen Prinzipien zugrunde liegen.

Die Cohensche "induktive Theorie" tritt nicht nur mit dem Anspruch auf, ein "wahres" Explikat des juristischen Rasonierens zu sein, sondern richtet sich darüber hinaus dagegen, daß erstens der mathematische Wahrscheinlichkeitsbegriff der einzige (formal explizierbare) Wahrscheinlichkeitsbegriff und zweitens das induktiv-probabilistische (alltägliche) Rasonieren - also nicht nur das juristische einzig und allein durch den mathematischen Begriff der Wahrscheinlichkeit präzisierbar ist. Es erscheint daher interessant, seine Einwände gegen die klassische Wahrscheinlichkeitstheorie näher zu studieren.

Um seine Position zu verstärken, stellt Cohen eine Reihe - wie er sie bezeichnet - Paradoxien auf, die seiner Meinung nach aus der Anwendung der Wahrscheinlichkeitstheorie auf die Rechtspraxis entstehen⁴⁾. Im folgenden werden wir diese Paradoxien untersuchen und

1) Allerdings liegen die deutschen Beiträge, die sich, wie bereits erwähnt, eher mit der Frage nach der Existenzberechtigung von Wahrscheinlichkeitsurteilen im juristischen Denken als mit dem Wahrscheinlichkeitsbegriff selbst bzw. seiner Anwendung auf die Rechtspraxis befassen, fast ausschließlich in Buchform vor.

2) Cohen meint, daß die Intuition, die dem mathematischen Wahrscheinlichkeitsbegriff zugrunde liegt, auf Pascal zurück geht; die Intuition, die seinen Prinzipien zugrunde liegt, führt er dagegen auf Bacon zurück; vgl. Cohen, S.2ff.

3) Mißverständlich, weil die Bezeichnung "induktiv" gewöhnlich in Verbindung mit dem herkömmlichen Wahrscheinlichkeitsbegriff gebracht wird.

4) Vgl. Cohen, S.2: "But, if one takes a careful enough look at some of the problems involved, quite large cracks begin to be discernible in the alleged closeness of fit between Pascalian calculus and forensic probability. Attempts can be made to paper over these cracks, but no such attempt could be fully successful. Paradox after Paradox emerges from any sufficiently thoroughgoing attempt to represent the logic of legal proof in terms of the calculus of chance".

zeigen, daß ihre Auflösung keine prinzipiellen Schwierigkeiten bereitet.

Was die "induktive" Theorie Cohens angeht, so läßt sich hier nur so viel sagen, daß sie gerade die Intuition als richtig hinstellt und formal zu rechtfertigen versucht, die ihrerseits den Anlaß zur Kritik und Beschäftigung mit der Wahrscheinlichkeitstheorie gab, weil sie die Unsicherheit, die im juristischen Rasonieren enthalten ist, nicht explizit berücksichtigt¹⁾.

Auf die oben genannten Bücher von Finkelstein und Eggleston wollen wir hier nicht eingehen, zumal das erstere sich an den früheren von Finkelstein publizierten Aufsätze orientiert (darunter auch der auf S.123ff dieser Arbeit besprochene Aufsatz, den er zusammen mit Fairley geschrieben hat) und das zweite ein einfaches und nichttechnisch geschriebenes Werk darstellt, das Juristen in die wahrscheinlichkeitstheoretische Problematik einführt.

1) Ausführlicher zu diesem Punkt vgl.S.72, 74ff dieser Arbeit.

Cohen

Cohen vertritt die Auffassung, daß der intuitive Wahrscheinlichkeitsbegriff, der den juristischen Überlegungen zugrunde liegt, nicht durch den mathematischen Begriff der Wahrscheinlichkeit (gleich, wie man ihn interpretiert) präzisiert werden kann, sondern durch einen ganz anders strukturierten Wahrscheinlichkeitsbegriff, für welchen weder das Negationsprinzip noch das Multiplikationsprinzip der Wahrscheinlichkeitstheorie gelten. Versucht man dagegen, den intuitiven juristischen Wahrscheinlichkeitsbegriff durch den mathematischen zu explizieren, so entstehen nach Cohen Paradoxien.

Wir wollen diese Paradoxien der Reihe nach mit ihren Lösungen dem Leser vorstellen.

1. Paradoxie: die Konjunktion¹⁾

Nach dem Multiplikationsprinzip²⁾ ist die Wahrscheinlichkeit für das gemeinsame Eintreten mehrerer Ereignisse kleiner (oder höchstens gleich, falls die restlichen Wahrscheinlichkeiten = 1 sind) als die Wahrscheinlichkeit für das Eintreten der jeweiligen Einzelereignisse, falls diese unabhängig voneinander sind. Dies ist nach Cohen für die Rechtspraxis unannehmbar: Besteht etwa ein Zivilprozeß aus dem Beweis mehrerer Behauptungen, von denen jede mit überwiegender Wahrscheinlichkeit zugunsten des Klägers festgestellt worden ist, so könnte es passieren, daß die gemeinsame Wahrscheinlichkeit für alle diese Behauptungen aufgrund des Multiplikationsprinzips nicht mehr überwiegt (z.B. wenn jeder der beiden Behauptungen mit einer Wahrscheinlichkeit von 0.7 festgestellt wurde, ist ihre gemeinsame Wahrscheinlichkeit nach diesem Prinzip = 0.49), so daß zugunsten des Angeklagten entschieden werden muß, was 'contra-intuitiv' ist. Zur Illustration führt Cohen folgende Fälle an³⁾

1. Angenommen, ein Autofahrer verklagt seine Versicherungsgesellschaft, weil diese es ablehnt, ihn nach einem Autounfall zu entschädigen. Weiter sei angenommen, daß zwei Sachverhalte nachgewiesen werden müssen: erstens, daß die Umstände des Unfalls zugunsten des Klägers sprechen und zweitens, daß der Wortlaut des Versicherungsvertrags des Klägers zugunsten des Klägers ausgelegt werden kann⁴⁾

Falls jeder dieser Sachverhalte mit einer Wahrscheinlichkeit von 0.71 (zugunsten des Klägers) nachgewiesen wird, ist ihre gemeinsame Wahrscheinlichkeit 0.504,

1) Cohens Bezeichnung: "The difficulty about Conjunction", S. 58ff.

2) Vgl. S. 16 dieser Arbeit.

3) Cohen, S. 59. Das gleiche Argument bringen in Anlehnung an Cohen Brilmayer & Kornhauser, S. 145f.

4) Das ist eine, meiner Ansicht nach, etwas präzisere Wiedergabe dessen, was Cohen mit der vagen Formulierung (S. 59) "Suppose, the two component issues that are disputed are first, what were the circumstances of the crash, and secondly, what were the terms of the driver's insurance contract" meint.

was nach dem Prinzip der überwiegenden Wahrscheinlichkeit für den Kläger ausreicht, um zu gewinnen. Sind jedoch die obigen Einzelwahrscheinlichkeiten z.B. nur 0.501, so ist die gemeinsame Wahrscheinlichkeit 0.251, und der Kläger muß verlieren.

2. Angenommen, eine zusammengesetzte Behauptung beschreibt eine Reihe voneinander unabhängiger Ereignisse. Ferner sei angenommen, daß für die Feststellung jedes einzelnen Ereignisses jeweils ein unabhängiger Zeuge (oder eine unabhängige Gruppe von Zeugen) erforderlich ist. Dann führt das dazu, daß die zusammengesetzte Behauptung um so unwahrscheinlicher wird, je größer die Anzahl der Zeugen ist. Vgl.S.60: "For example, if a series of independent points are all conjoined in a single allegation, and to establish each point a separate witness (or a group of witnesses) is needed, then the higher the number of witnesses (or groups of witnesses) that is needed, the more reliable each witness (or group of witnesses) has to be. So if no witness (or group of witnesses) were ever to have a more than .9 probability of speaking the truth, no case could ever deserve to win which conjoined more than six component points that were mutually independent and each required a separate witness (or group of witnesses) - since .9⁷ is less than .5".

Wir beginnen die Analyse mit dem ersten Fall:

Hier haben wir es mit einer komplexen Sachverhaltshypothese (wir kürzen sie mit H ab) zu tun, die aus zwei einfachen Sachverhaltshypothesen (wir kürzen sie mit H_1 für "die Umstände des Unfalls sprechen zugunsten des Klägers" und H_2 für "Der Wortlaut des Versicherungsvertrags kann zugunsten des Klägers ausgelegt werden" ab) zusammengesetzt ist¹⁾

Formal rekonstruiert haben wir es im Fall der einfachen Hypothese H_1 ($i=1,2$) mit einem Hypothesenraum²⁾ zu tun, der nur aus zwei möglichen Hypothesen besteht:

H_1 und $\neg H_1$ (Elemente des Hypothesenraums der Hypothese H_1);

H_2 und $\neg H_2$ (Elemente des Hypothesenraums der Hypothese H_2).

Im Fall der zusammengesetzten Hypothese H haben wir es dagegen mit einem Hypothesenraum zu tun, der aus vier Elementen besteht. Diese sind:

$H_1 \wedge H_2$, $H_1 \wedge \neg H_2$, $\neg H_1 \wedge H_2$, $\neg H_1 \wedge \neg H_2$ (Elemente des Hypothesenraums der Hypothese H).

1) Es ist fraglich, inwiefern man H_2 , die die Auslegung eines juristischen Textes (und somit keine empirische, sondern eine Rechtsfrage) betrifft, wahrscheinlichkeits-theoretisch behandeln kann. Wir wollen auf die Klärung dieser Frage hier nicht eingehen, zumal Cohen hier die grundsätzliche Problematik anspricht, daß eine juristische Sachverhaltshypothese zusammengesetzt werden kann, und weil wir darüber hinaus genug Fälle finden können, in denen es sich bei den einzelnen Hypothesen, aus denen die zusammengesetzte Sachverhaltshypothese besteht, um empirische Hypothesen handelt.

2) Zum Begriff des "Möglichkeitsraumes" (oder, wie hier, des "Hypothesenraumes") vgl. S. 11 dieser Arbeit.

Ordnen wir sowohl der Hypothese H_1 wie auch der Hypothese H_2 (die nach Voraussetzung voneinander unabhängig sind) den von Cohen angenommenen Wahrscheinlichkeitswert 0.501 zu, so erhalten wir:

$$\begin{aligned} P(H_1) &= 0.501 \rightarrow P(\neg H_1) = 0.499; P(H_1) > P(\neg H_1), \text{ somit } \text{überwieg}t H_1; \\ P(H_2) &= 0.501 \rightarrow P(\neg H_2) = 0.499; P(H_2) > P(\neg H_2), \text{ somit } \text{überwieg}t H_2; \\ P(H_1 \wedge H_2) &= 0.251, P(H_1 \wedge \neg H_2) = 0.250, P(\neg H_1 \wedge H_2) = 0.250, \\ P(\neg H_1 \wedge \neg H_2) &= 0.249. \\ P(H_1 \wedge H_2) &> P(H_1 \wedge \neg H_2); > P(\neg H_1 \wedge H_2); > P(\neg H_1 \wedge \neg H_2). \end{aligned}$$

Wie wir sehen, überwiegt $H_1 \wedge H_2$ auch im zusammengesetzten Fall, denn $H_1 \wedge H_2$ ist unter allen möglichen zusammengesetzten Hypothesen aus dem Hypothesenraum am wahrscheinlichsten. Entscheidet das Gericht im Falle, daß jede Sachverhaltshypothese mit der Wahrscheinlichkeit von 0.501 zugunsten des Klägers nachgewiesen ist, im Endurteil auch zugunsten des Klägers, so ist das wie folgt wahrscheinlichkeits-theoretisch zu interpretieren:

Das Gericht befolgt offensichtlich hier nicht die Regel, wonach $P(H_1 \wedge H_2) > 0.5$ (oder wonach der zusammengesetzte strafbare Sachverhalt wahrscheinlicher als seine zusammengesetzte Negation, d.h. alle diejenigen Fälle, in welchen mindestens eine der Komponenten nicht gilt) sein soll, sondern die Regel, wonach $H_1 \wedge H_2$ lediglich wahrscheinlicher als jede andere Möglichkeit (bzw. jeder andere in Frage kommende Sachverhalt) sein muß¹⁾

Der Mißgriff Cohens besteht also darin, daß er zwar $P(H_1 \wedge H_2)$ gegen $P(\neg(H_1 \wedge H_2))$ abwägt (was an sich zulässig ist), wobei er annimmt, daß $P(H_1 \wedge H_2) > P(\neg(H_1 \wedge H_2))$ sein soll (was plausibel ist), um der Forderung "mit überwiegender Wahrscheinlichkeit" zu genügen, jedoch (ungewollt) ein anderes Prinzip anspricht, wenn er schreibt, daß das Gericht auch dann zugunsten des Klägers entscheiden soll²⁾, wenn zwar jede einzelne Hypothese ihre Negations überwiegt, $P(H_1 \wedge H_2)$ jedoch kleiner als $P(\neg(H_1 \wedge H_2))$ ist³⁾. Aus wahrscheinlichkeits-theo-

1) Die Natur des obigen Zusammenhangs ist etwas komplexer und es ist uns nicht möglich, an dieser Stelle darauf näher einzugehen. Im II. Teil werden wir ein Prinzip angeben, das den Wahrscheinlichkeitswert für den Sachverhalt festlegt, der für die entscheidungstheoretische Beurteilung des Falls ausreicht.

2) Es ist allerdings nicht erwiesen, daß das Gericht sich in der Tat so verhält, und es soll hier dahin gestellt bleiben, ob eine solche (von Cohen eruierte) Entscheidung vernünftig ist.

3) Cohen meint, daß die Wahrscheinlichkeitstheorie das Gericht zwingt nach dem unwahrscheinlicheren Sachverhalt zu entscheiden, falls $P(H_1 \wedge H_2) < P(\neg(H_1 \wedge H_2))$ ist.

retischer Sicht wird man sagen, daß hier das Gericht nach dem wahrscheinlicheren Sachverhalt vorgegangen ist (im oben besprochenen Beispiel überwiegt $H_1 \wedge H_2$ jeden anderen möglichen zusammengesetzten Sachverhalt)¹⁾ Das Ergebnis, daß das Gericht zugunsten des Klägers entscheidet, falls jeder Sachverhalt mit überwiegender Wahrscheinlichkeit nachgewiesen ist (gleichgültig, wie groß dieses Überwiegen sein mag), ist somit mit der Wahrscheinlichkeitstheorie konsistent²⁾³⁾

Schwieriger ist die Interpretation des zweiten Falls; denn es ist nicht ganz klar, was Cohen mit der Formulierung "if a series of independent points are all conjoined in a single allegation" konkret meint: Dies können nämlich sowohl mehrere Beweismittel oder aber auch eine zusammengesetzte Sachverhaltshypothese sein. Wir untersuchen die beiden Möglichkeiten:

1. Falls Cohen mit der obigen Formulierung sich auf mehrere (voneinander unabhängige) Beweismittel bezieht, so ist die Anwendung des Multiplikationsprinzips bei der Bestimmung der Wahrscheinlichkeit für die Sachverhaltshypothese aufgrund dieser Beweismittel nicht korrekt.

1) "überwiegt" ist hier also nicht im Sinne Cohens als "überwiegt mit der Wahrscheinlichkeit größer 0.5" zu verstehen, sondern als "wahrscheinlicher als jeder andere mögliche Sachverhalt". Auch im folgenden wird "überwiegt" in diesem letzteren Sinn verstanden. Für das "Überwiegen" im ersteren Sinn werden wir hingegen die Bezeichnung "überwiegt im Sinn > 0.5 " oder, einfach, "überwiegt im Cohenschen Sinn" verwenden.

2) Besteht das Gericht darauf, daß auch $P(H_1 \wedge H_2) > 0.5$ sein soll, so geht es dann offensichtlich im Fall der einzelnen Hypothese nach einer strengeren Regel vor als bloß nach "überwiegt im Sinn von > 0.5 ".

3) Es ist eine andere Frage, ob eine solche Entscheidung auch mit der Entscheidungstheorie konsistent ist. Denn es ist fraglich, inwiefern die Wahrscheinlichkeitstheorie (wie ihr von Cohen unterstellt wird) allein bereits eine Lösung des Falls, d.h. die Entscheidung bzgl. der Verurteilung oder des Freispruchs, liefern kann. Rational gesehen, müssen juristische Fälle entscheidungstheoretisch rekonstruiert werden. Im entscheidungstheoretischen Rahmen werden auch bestimmte Beweiskriterien, die gewisse, plausibel erscheinende Anforderungen an den Wahrscheinlichkeitswert für den Sachverhalt stellen, formuliert. Vgl. den II. Teil dieser Arbeit.

Daß Cohen der Wahrscheinlichkeitstheorie unterstellt, bereits eine Lösung des Falls liefern zu können, belegt die folgende Stelle (S.78) "On the mathematicist account a man might apparantly lose his whole fortune if those of his circumstances that were in evidence tended to place him on the majority side of some critical division even though he actually belonged with the minority".

Wir wollen die (unter 2.S.173 dieser Arbeit) etwas allgemeine Ausführung Cohens an einem Beispiel veranschaulichen. Wir nehmen an, daß zwei voneinander unabhängige Beweismittel im Prozeß vorgelegt werden: (1) daß der Angeklagte wegen Totschlags ein Motiv hat (Abk. M), was durch die Aussage des Zeugen X bestätigt wird (Abk. A)¹⁾ und (2) daß der Angeklagte am Tatort war (Abk. T), die Aussage eines anderen Zeugen Y bestätigt das (Abk. B)²⁾

Wir kürzen die bedingte Wahrscheinlichkeit, daß der Angeklagte ein Motiv hat aufgrund der Zeugenaussage von X, mit $P(M/A)$ ab und die bedingte Wahrscheinlichkeit, daß der Angeklagte am Tatort war aufgrund der Zeugenaussage von Y, mit $P(T/B)$.

Cohen will vermutlich³⁾ die Wahrscheinlichkeit für das gemeinsame Vorliegen beider Beweismittel aufgrund der beiden (voneinander unabhängigen) Aussagen der Zeugen X und Y (Abk. $P(M \wedge T/A \wedge B)$) feststellen, wobei nach Cohen offensichtlich gilt: $P(M \wedge T/A \wedge B) = P(M/A)P(T/B)$ (vgl. den Wortlaut des Zitats⁴⁾). Cohen benutzt also das Multiplikationsprinzip und begeht damit genau denselben Fehler, der bereits im Fall 'People v. Collins' so verhängnisvoll war: denn die aufgrund des Multiplikationsprinzips festgestellte Wahrscheinlichkeit besagt lediglich, daß das gemeinsame Vorliegen mehrerer Beweismittel weniger wahrscheinlich ist als das Vorliegen jedes einzelnen Beweismittels (ihre Unabhängigkeit vorausgesetzt).

1) X könnte z.B. aussagen: "Der Angeklagte hat mir gegenüber gestanden, er habe den Verstorbenen gehaßt".

2) Y könnte z.B. behaupten, er habe den Angeklagten (den er gut kennt) am Tatort gesehen.

3) Denn es ist nicht ganz klar, wie Cohens Formulierung ".9 probability of speaking the truth" zu verstehen ist; dies kann sich nämlich auch auf die Zeugenglaubwürdigkeit beziehen. Da Cohen aber die Wahrscheinlichkeit von 0.9⁷ für das "speaking the truth" der 7 Zeugen mit der Wahrscheinlichkeit für die zusammengesetzte Behauptung identifiziert (vgl. den Schluß des Zitats), ist anzunehmen, daß er mit diesen Wahrscheinlichkeiten etwas ähnliches wie $P(B_1/Z_1), \dots, P(B_n/Z_n)$ bzw. im zusammengesetzten Fall: $P(B_1 \wedge \dots \wedge B_n/Z_1 \wedge \dots \wedge Z_n)$ meint ($B_1, \dots, B_n$ sind einzelne Behauptungen, $Z_1, \dots, Z_n$ einzelne Zeugenaussagen).

4) Dies ist allerdings dann richtig, wenn M und T voneinander (bedingt auf $A \wedge B$) unabhängig sind (und außerdem, wenn B für M und A für T irrelevant sind, was nach Voraussetzung gilt). Es ist falsch von der Unabhängigkeit der Zeugenaussagen auf die Unabhängigkeit der Beweismittel, welche diese Zeugenaussagen stützen, zu schließen (wie Cohen dies nach der obigen Interpretation tut). Gerade im vorliegenden Fall, wo die zwei Zeugenaussagen A und B unabhängig voneinander sind, gilt das für die beiden Beweismittel offensichtlich nicht, denn es ist sehr selten, daß man zwar ein Motiv besitzt und am Tatort war, dies jedoch miteinander nichts zu tun hat.

Keineswegs kann man daraus auf die Täterschaft bzw. Schuld des Angeklagten schließen¹⁾ Im Prozeß interessiert uns aber immer das letztere und um das festzustellen, müssen wir das Bayes-Theorem verwenden, indem wir die Wahrscheinlichkeit der Sachverhaltshypothese, daß der Angeklagte die Tat begangen hat, aufgrund der beiden Beweismittel "Motiv" und "Anwesenheit am Tatort" bestimmen. Die Aposteriori-Wahrscheinlichkeit der Hypothese aufgrund mehrerer Beweismittel hängt dann nicht nur von der Apriori-Wahrscheinlichkeit der Hypothese, sondern auch von der Zuverlässigkeit (bzw. der Glaubwürdigkeit) einzelner Beweismittel ab. Somit kann der Fall eintreten, daß die Aposteriori-Wahrscheinlichkeit sehr hoch ist, wenn die Zuverlässigkeit der Beweismittel hoch ist; und umgekehrt kann die Aposteriori-Wahrscheinlichkeit klein sein, falls die Beweismittel nicht zuverlässig sind. Dies gibt wahrscheinlichkeitstheoretisch die Intuition wieder, die auch im juristischen Prozeß geläufig ist.

Anmerkungen:

1. Im vorliegenden Fall müssen wir allerdings die Wahrscheinlichkeit der Sachverhaltshypothese bzgl. der Täterschaft des Angeklagten nicht mit Hilfe des Bayes-Theorems, sondern mit der B_{erw} -Formel (vgl.S. 34 dieser Arbeit) bestimmen. Es sei daran erinnert, daß die B_{erw} -Formel eine Erweiterung der Bayes-Formel ist, mit welcher man die Unsicherheit hinsichtlich der der Hypothese zugrundeliegenden Informationen (im vorliegenden Fall: "Motiv" und "Anwesenheit am Tatort") dadurch eliminieren kann, daß man die Wahrscheinlichkeit der Hypothese H , statt in bezug auf M und T , relativ zu den sicheren Informationen (hier sind es die beiden Zeugenaussagen) beurteilt. Die Formel lautet:

$$P(H/A \wedge B) = P(H/(M \wedge T) \wedge A \wedge B)P(M \wedge T/A \wedge B) + P(H/-(M \wedge T) \wedge A \wedge B)P(-(M \wedge T)/A \wedge B).$$

Wäre die Unsicherheit hinsichtlich der Beweismittel im vorliegenden Fall nicht auf diese Weise eliminierbar, so müßten wir hier die Jeffrey-Formel anwenden:

$$P(H/P*(M \wedge T)) = P(H/M \wedge T)P*(M \wedge T) + P(H/-(M \wedge T))P*(-(M \wedge T)).$$

2. Es kann allerdings die Situation eintreten, daß mehrere Beweismittel vorliegen, deren jedes mit einer hohen Wahrscheinlichkeit festgestellt ist und deren jedes die Hypothese (sogar!) impliziert und daß trotzdem die Wahrscheinlichkeit der Hypothese relativ zu diesen Beweismitteln nicht größer ist als die Wahrscheinlichkeit für das gemeinsame Vorliegen der Beweismittel. Ein Ergebnis also, welches Cohen 'contra-intuitiv' nennen würde.

Schauen wir uns die Sachlage näher an:

Nehmen wir an, daß es sich dabei (wie in Cohens Fall, S.173 dieser Arbeit) um 7 voneinander unabhängige Zeugenaussagen handelt, deren jede mit der Wahrscheinlichkeit von 0.9 festgestellt ist (wir kürzen die einzelnen Zeugenaussagen mit $D_1, \dots, D_7$ ab). Es gilt also: $P*(D_1) = \dots = P*(D_7) = 0.9$. Die Wahrscheinlichkeit für das gemeinsame Vorliegen dieser Beweismittel, d.h. $P*(D_1 \wedge \dots \wedge D_7)$, ist daher ca. 0.48.

1) Zu diesem Punkt siehe die Diskussion um den Collins-Fall, S.112ff dieser Arbeit.

Ferner nehmen wir an, daß gilt: $P(H/D_1 \wedge \dots \wedge D_7) = 1$, d.h. die Zeugenaussagen implizieren die Hypothese. Ist $P(H/\neg(D_1 \wedge \dots \wedge D_7)) = 0$ (d.h. die Negation der Beweismittel impliziert die Falschheit der Hypothese), dann folgt mit der Jeffrey-Formel¹⁾

$P(H//P*(D_1 \wedge \dots \wedge D_7)) = 0.48$; außerdem erhalten wir:
 $P(\neg H//P*(D_1 \wedge \dots \wedge D_7)) = 0.52$; d.h. auch dann, wenn jedes Beweismittel mit einer hohen Wahrscheinlichkeit festgestellt ist und außerdem die Hypothese sicher macht, ist die Wahrscheinlichkeit der Hypothese aufgrund dieser Beweismittel kleiner als eine solche für die Negation dieser Hypothese aufgrund der gleichen Beweismittel.

Etwas genauer: Seien $P(D_1 \wedge \dots \wedge D_7/H) = 1$ und $P(D_1 \wedge \dots \wedge D_7/\neg H) = 0$; eingesetzt in die Bayes-Formel erhalten wir: $P(H/D_1 \wedge \dots \wedge D_7) = 1$ und somit $P(\neg H/D_1 \wedge \dots \wedge D_7) = 0$. Außerdem, da $P(D_1 \wedge \dots \wedge D_7/H) = 1 + P(\neg(D_1 \wedge \dots \wedge D_7)/H) = 0$ und $P(D_1 \wedge \dots \wedge D_7/\neg H) = 0 \rightarrow P(\neg(D_1 \wedge \dots \wedge D_7)/\neg H) = 1$ folgt $P(\neg H/\neg(D_1 \wedge \dots \wedge D_7)) = 1$.

Eingesetzt in die Jeffrey-Formel erhalten wir:

$$P(H//P*(D_1 \wedge \dots \wedge D_7)) = 1 \cdot 0.48 + 0 \cdot 0.52 = 0.48;$$

$$P(\neg H//P*(D_1 \wedge \dots \wedge D_7)) = 0 \cdot 0.48 + 1 \cdot 0.52 = 0.52.^{2)}$$

Im obigen Beispiel haben wir allerdings die Voraussetzung gemacht, daß, wenn bereits ein einziges Beweismittel falsch ist (d.h. wenn etwa ein einziger Zeuge die Unwahrheit sagt), die Wahrscheinlichkeit für die Hypothese aufgrund aller Beweismittel = 0 ist; d.h. die falsche Aussage eines einzigen Zeugen genügt, um die Negation der Hypothese wahr zu machen³⁾. Unter dieser extrem unplausiblen Voraussetzung gewinnt selbstverständlich der zweite Teil der Jeffrey-Formel das Übergewicht, d.h. die Wahrscheinlichkeit der Negation der Hypothese unter der Bedingung, daß wenigstens ein Beweismittel falsch ist (diese ist im obigen Fall = 1), multipliziert mit der Wahrscheinlichkeit für die Falschheit von wenigstens einem Beweismittels (diese ist = 0.52); so daß letztlich die Negation der Hypothese aufgrund der unsicheren Daten insgesamt wahrscheinlicher ist als die Hypothese aufgrund dieser Daten (auch dann, wenn diese mit einer hohen Wahrscheinlichkeit festgestellt sind).

1) Die Jeffrey-Formel lautet in diesem Fall:

$$P(H//P*(D_1 \wedge \dots \wedge D_7)) = P(H/D_1 \wedge \dots \wedge D_7) \cdot P*(D_1 \wedge \dots \wedge D_7) + P(H/\neg(D_1 \wedge \dots \wedge D_7)) \cdot P*(\neg(D_1 \wedge \dots \wedge D_7))$$

$$\text{und entsprechend für die Negation: } P(\neg H//P*(D_1 \wedge \dots \wedge D_7)) = P(\neg H/D_1 \wedge \dots \wedge D_7) \cdot P*(D_1 \wedge \dots \wedge D_7) + P(\neg H/\neg(D_1 \wedge \dots \wedge D_7)) \cdot P*(\neg(D_1 \wedge \dots \wedge D_7)).$$

2) Falls jedes Beweismittel mit der Wahrscheinlichkeit von 0.99 festgestellt ist, 'genügen' 70 Beweismittel, um ein solches 'contraintuitives' Resultat zu erzielen (denn $0.99^{70} \approx 0.495$; somit erhält man im o.Beispiel: $P(H//P*(D_1 \wedge \dots \wedge D_{70})) < P(\neg H//P*(D_1 \wedge \dots \wedge D_{70}))$).

3) Denn es gilt: $\neg(D_1 \wedge \dots \wedge D_7) = (\neg D_1 \vee \neg D_2 \vee \dots \vee \neg D_7)$ v $(D_1 \wedge \neg D_2 \wedge \dots \wedge \neg D_7)$ v $\dots$ v $(\neg D_1 \wedge \dots \wedge \neg D_7)$.

$\neg(D_1 \wedge \dots \wedge D_7)$ besagt, daß wenigstens ein Beweismittel falsch ist.

3. Vieles spricht dafür, daß Cohen sein "Überwiegensprinzip" auf die Beweismittel anwendet.

Ist es jedoch (unter gewissen Einschränkungen) gerechtfertigt, in bezug auf die Sachverhaltshypothese von einem Überwiegensprinzip auszugehen (d.h. einem Prinzip, wonach der Fall zugunsten des Klägers entschieden wird, wenn jede Komponente der zusammengesetzten Sachverhaltshypothese mit "überwiegender Wahrscheinlichkeit" zu seinen Gunsten festgestellt ist), so ist keineswegs selbstverständlich zu erwarten, daß dieses Prinzip auch in bezug auf die Beweismittel gelten soll (Hier besagt das "Überwiegensprinzip", daß der Fall zugunsten des Klägers entschieden wird, wenn bereits jedes der Beweismittel (d.h. der die Sachverhaltshypothese stützenden Daten), keineswegs die Sachverhaltshypothese selbst, mit "überwiegender Wahrscheinlichkeit" zugunsten des Klägers festgestellt ist).

Das Akzeptieren des "Überwiegensprinzips" in bezug auf die Beweismittel widerspricht übrigens auch der Behauptung Cohens¹⁾, wonach die Gerichtspraxis mehrstufige Wahrscheinlichkeitsurteile nicht zuläßt²⁾. Denn im vorliegenden Fall handelt es sich gerade um ein solches mehrstufiges Wahrscheinlichkeitsurteil: Es gilt nämlich sowohl ein Wahrscheinlichkeitsschluß von den Beweismitteln auf die Sachverhaltshypothese wie auch ein solcher von den weiteren Tatsachen (den beiden Zeugenaussagen) auf die Beweismittel.

Dafür, daß Cohen das "Überwiegensprinzip" eher auf die Beweismittel als auf die Sachverhaltshypothese anwendet, spricht die Tatsache, daß er seine "induktive Wahrscheinlichkeit" im Sinn von "weight of evidence" konzipiert. Vgl. auch S.36 wo er - um zu verdeutlichen, was er mit seinem Wahrscheinlichkeitsbegriff meint - Keynes (A Treatise on Probability (1957), S.71ff) zitiert: "One argument has more weight than another if it is based on a greater amount of relevant evidence..." (Unterstreichung von mir)³⁾

Es bleibt uns hier noch die zweite Interpretationsmöglichkeit der mehrdeutigen Formulierung Cohens "if a series of independent points are all conjoined in a single allegation..." und damit des zweiten Beispiels (S.173) zu erwähnen. Gegenüber der Analyse der ersten Interpretationsmöglichkeit fällt diese jedoch kurz aus.

2. Cohen könnte sich mit dieser Formulierung, statt auf zusammengesetzte Beweismittel, auf eine zusammengesetzte Sachverhaltshypothese beziehen. Falls dies stimmt, ist die Argumentation wie im ersten Fall (das erste Beispiel, S.172 dieser Arbeit) zu rekonstruieren⁴⁾

1) Cohen, S.68f; vgl. auch S.182ff dieser Arbeit.

2) Vgl. *ibid*, S.68: Where a proof in a civil case involves several tiers of inference, the courts normally insist that each tier prior to the final one should rest on proof beyond reasonable doubt". Wenn das stimmt, dürfen also solche Fälle wie der von mir auf S.176 konstruierte vor Gericht nicht zugelassen werden; d.h. alle Beweismittel, die auf Zeugenaussagen beruhen, müssen mit "fast-Sicherheit"(!) nachgewiesen werden.

3) Zu diesem Punkt vgl. auch S. 76f dieser Arbeit.

4) In der Rechtspraxis haben wir es stets mit zusammengesetzten Hy-

Zusammenfassung

Die von Cohen aufgestellte Paradoxie der Konjunktion löst sich auf. Das Vorgehen des Gerichts, auch dann zugunsten des Klägers zu entscheiden, wenn zwar jeder Sachverhalt mit überwiegender Wahrscheinlichkeit zu seinen Gunsten festgestellt wurde, die Wahrscheinlichkeit für die zusammengesetzte Sachverhaltshypothese jedoch kleiner als 0.5 ist, ist mit der Wahrscheinlichkeitstheorie konsistent. Denn in diesem Fall entscheidet sich das Gericht für den Sachverhalt, der wahrscheinlicher ist als jeder andere Sachverhalt aus dem entsprechenden Möglichkeitsraum.

Es sei hier allerdings noch einmal betont: erstens ist es nicht erwiesen, daß das Gericht in der Praxis nach der von Cohen aufgestellten Regel entscheidet; zweitens: würde sich Gericht in der Praxis doch nach dieser Regel orientieren, so können wir eine solche Entscheidungspraxis zwar mit der Forderung in Einklang bringen, daß die Entscheidung entsprechend dem wahrscheinlicheren (zusammengesetzten) Sachverhalt aus dem Möglichkeitsraum gewählt werden soll. Ob eine solche Entscheidung auch entscheidungstheoretisch konsistent (rational und dergl.) bleibt, ist fraglich¹⁾ und läßt sich im Rahmen der Wahrscheinlichkeitstheorie nicht beantworten.

Die Anwendung des Multiplikationsprinzips ist nur dann zulässig, wenn es sich um eine zusammengesetzte Sachverhaltshypothese handelt. Handelt es sich dagegen um mehrere Beweismittel, die ihrerseits zur Beurteilung der Sachverhaltshypothese herangezogen werden, so muß man hier, will man korrekt vorgehen, die Bayes- (bzw. Jeffrey-. bzw. Bernoulli-) Formel, auf keinen Fall aber das Multiplikationsprinzip anwenden²⁾.

Die Fehler der Cohenschen Argumentation sind somit folgende:

1. Eine unkorrekte Rekonstruktion des Hypothesenraumes der in Frage kommenden Sachverhaltshypothese (dieser muß vollständig sein und aus paarweise disjunkten Ereignissen bestehen) und somit die fehlende Berücksichtigung der Tatsache, daß das Gericht im Einklang mit

pothesen zu tun bei der Beurteilung von sog. "qualifizierten" Tatbeständen wie Mord, schwerer Diebstahl, Raub usw., denn in solchen Fällen müssen wir nicht nur nachweisen, daß der Angeklagte das Opfer getötet, bestohlen usw. sondern auch, daß er das z.B. auf eine grausame Art und Weise (Mord) oder mit Anwendung von Gewalt (Raub) etc. getan hat. In diesen Fällen müssen wir das Multiplikationsprinzip anwenden.

- 1) Im Beispiel auf S.174 folgt auf drei von vier möglichen zusammengesetzten Sachverhalte (nach der Regel, daß jede Komponente der zusammengesetzten Sachverhalt überwiegen muß) Freispruch. Außerdem ist die Wahrscheinlichkeit dafür, daß einer dieser drei Sachverhalte vorliegt, größer als 0.5. Intuitiv gesehen würde man eher geneigt sein, hier den Freispruch (und nicht die Verurteilung) zu wählen.
- 2) Daß Cohen im zweiten Beispiel das Multiplikationsprinzip auf die Beweismittel anwendet, bestätigt u.a. ihre große Zahl (vgl. 7 im Zitat auf S.173 dieser Arbeit), die er dabei annimmt. Keine zusammengesetzte juristische Sachverhaltshypothese wird meines Erachtens jemals so viele Einzelglieder aufweisen. Die Anzahl der Beweismittel kann dagegen sogar in die Hunderte gehen. (Vgl.dazu auch S.179 dieser Arbeit

der Wahrscheinlichkeitstheorie nach dem wahrscheinlicheren Sachverhalt entscheidet, falls es (nach der von Cohen postulierten Regel) die Entscheidung entsprechend dem zusammengesetzten Sachverhalt wählt, wenn jede Komponente dieses Sachverhalts ihrerseits mit überwiegender Wahrscheinlichkeit im Sinn > 0.5 festgestellt ist.

2. Eine fehlende Unterscheidung zwischen einer zusammengesetzten Sachverhaltshypothese und einem zusammengesetzten Datum (mehrere Beweismittel) und somit

3. Eine unzulässige Anwendung des Multiplikationsprinzips in Fällen, wo das Bayes-Theorem bzw. die Jeffrey-, bzw. Berw-Formel anzuwenden ist¹⁾ bzw.

4. eine unzulässige Erweiterung des Gültigkeitsbereichs des "Überwiegensprinzips" von einer Sachverhaltshypothese auf Beweismittel.

1) Das Bayes-Theorem scheint Cohen wenig zu interessieren, denn er erwähnt es mit keinem Wort. Zur Kritik an Cohen, u.a. daß er das Bayes-Theorem nicht berücksichtigt, vgl. G.Williams, The Mathematics of Proof, in: Criminal Law Rev. (1979), S.297-308, 340-354.

2. Paradoxie: die fortgesetzte Schlußfolgerung¹⁾

Diese Paradoxie entsteht nach Cohen immer dann, wenn man die Wahrscheinlichkeitstheorie auf juristische Fälle anwendet, die aus mehreren Beweisstufen bestehen, so daß eine Stufe zum Nachweis einer anderen und diese wiederum zum Nachweis einer dritten (etwa wenn Q zum Nachweis von R und R zum Nachweis von S) usw. benutzt wird. Das Gericht besteht nach Cohen in solchen Fällen (auch im Zivilprozeß) darauf, daß alle Stufen, die der letzten vorangehen mit der "an Sicherheit grenzenden Wahrscheinlichkeit" festgestellt werden²⁾. Der Einwand Cohens gegen die Wahrscheinlichkeitstheorie ist, daß diese auch dann zur Ableitung eines Endergebnisses nach einem juristischen Beweiskriterium (etwas wie im Zivilprozeß "mit überwiegender Wahrscheinlichkeit") erlaubt, wenn jede Stufe (also nicht nur die letzte) mit 'nur' überwiegender Wahrscheinlichkeit nachgewiesen ist.

Der obige Einwand Cohens scheint auf einigen Begriffsunklarheiten und Mißverständnissen zu beruhen. Wir werden im folgenden untersuchen, woran das liegen mag:

1. Bevor man den Ausdruck "mit an Sicherheit grenzender Wahrscheinlichkeit", den das Gericht (nach Cohens Auffassung) für jede Stufe (ausgenommen der letzten im Zivilprozeß) eines mehrstufigen Beweises verlangt, im Rahmen einer Kritik verwendet, soll man seine Bedeutung klären, was Cohen nicht tut³⁾. Angesichts dessen, daß ver-

1) Cohen bezeichnet diese Paradoxie als "the difficulty about inference upon inference", S. 68.

2) *ibid.*, S. 68ff.

3) Zum Begriff mit "an Sicherheit grenzender Wahrscheinlichkeit" vertritt Cohen eine konservative und wenig klare Auffassung; vgl. S. 249: "...if the defence challenges the prosecution's argument at every possible point, what the prosecution has to do, in effect, is to establish that every relevant feature of the situation is one of those that are favourable, or not abnormal, for the generalization invoked: e.g. that no other person was in the vicinity, that the accused was in full possession of his senses and knew what he was doing, etc., etc. In a relatively uncomplicated case proof beyond reasonable doubt - i.e. an inductive probability that amounts to virtual certainty - is achieved when every let-out of this nature is eliminated, either by oral, documentary, or other evidence, or by reference to facts that the defence admits or the court is prepared to notice". Vgl. dazu auch die Kritik von Schum, S. 482: "Within P_I (die Cohensche induktive Wahrscheinlichkeit - Anm. von mir), proof beyond a reasonable doubt means proof at a level of inductive certainty; i.e., when there is no specific reason for doubting the truth of a conclusion. Maximum $P_I(H,E)$ is obtained when evidence E includes all relevant points to be established together with satisfaction that the prosecution has proven these points. I, for one, have difficulty imaging a situation in which all relevant points have been listed, let alone established in fact".

schiedene Explikate dieses Begriffs auch zu unterschiedlichen Schlußfolgerungen sowohl bzgl. der Anwendung der Wahrscheinlichkeitstheorie auf die Rechtspraxis wie auch der Rechtspraxis selbst führen, wäre eine solche Explikation jedoch unumgänglich. Denn man kann diesen Begriff als ein Beweiskriterium auffassen, das nicht objektiv ist und von Individuum zu Individuum und von Fall zu Fall variiert. Der Unterschied zwischen diesem subjektiven Beweiskriterium und dem der "überwiegenden Wahrscheinlichkeit" fällt - wie empirisch belegt - gering aus (Jeder Wert größer als 0.5 kann dieses Beweiskriterium repräsentieren¹⁾). Daher haben wir vorgeschlagen (S.169), anstelle dieses Ausdrucks von der Wahrscheinlichkeit zu sprechen, die für ein Individuum in einem konkreten Fall für die Verurteilung (bzw. für eine bestimmte Schlußfolgerung, einen bestimmten Beweis) ausreicht).

Akzeptiert man, daß - wie die empirische Untersuchungen ergaben - für ein Individuum ein beliebiger Wert größer als 0.5 die Wahrscheinlichkeit repräsentieren kann, die für einen weiteren Beweis ausreicht, so ist es schwer einzusehen, wieso die Wahrscheinlichkeitstheorie auf juristische Fälle nicht anwendbar sein soll, die mehrere Beweisstufen enthalten.

Faßt man dagegen den Ausdruck "mit an Sicherheit grenzender Wahrscheinlichkeit" in dem Sinne auf, daß "nur erkenntnistheoretische Zweifel, die immer bestehen", dadurch zum Ausdruck gebracht werden sollen²⁾ - wie Cohen es offensichtlich meint³⁾ - so hat das für die Rechtspraxis die Konsequenz, daß alle juristischen Fälle, die auf der Unsicherheit von Beweismitteln (wie z.B. fast alle Zeugenbeweise) beruhen, zugunsten des Angeklagten entschieden werden müssen⁴⁾.

Die Zeugenbeweise haben i.d.R. die Struktur von mehrstufigen Beweisen, wie etwa in dem auf S.176 dieser Arbeit geschilderten Beispiel, in dem die Aussage des Zeugen X zum (wahrscheinlichkeitstheoretischen) Nachweis dient, daß der Angeklagte ein Motiv hat,

1) Vgl.S.169 dieser Arbeit.

2) Vgl.S.58 Fn.1 dieser Arbeit.

3) vgl.S.182 Fn.3 dieser Arbeit.

4) Vgl.S.179 Fn.2 dieser Arbeit.

und das wiederum (d.h. daß der Angeklagte das Motiv hat) zum (wahrscheinlichkeitstheoretischen) Nachweis, daß der Angeklagte der Täter ist. Was Cohen also in einem solchen Fall verlangen würde, ist, daß der Zeuge das Motiv des Angeklagten "mit an Sicherheit grenzender Wahrscheinlichkeit" (wobei dieser Ausdruck in dem von Cohen intendierten Sinn zu verstehen ist) 'beweist'. Angesichts dessen, daß Zeugenbeweise überhaupt die problematischsten juristischen Beweise sind und daher kaum ein Richter zu finden sein wird, der "nur erkenntnistheoretische Zweifel" am Aussagewert einer Zeugenaussage haben würde, kann Cohens Forderung¹⁾ vielleicht theoretisch in Erwägung gezogen werden, kaum aber die gewöhnliche Vorgehensweise in der Rechtspraxis repräsentieren²⁾

2. Aber auch dann, wenn man die Bedeutung des Ausdrucks "mit an Sicherheit grenzender Wahrscheinlichkeit" geklärt hat, kann man fragen, inwiefern die von Cohen angesprochene Thematik die Wahrscheinlichkeitstheorie überhaupt tangiert. Denn ob das Gericht einen Wahrscheinlichkeitswert z.B. von 0.7 für den weiteren Beweis für ausreichend hält oder nicht, kann vielleicht juristisch oder entscheidungstheoretisch gelöst werden, keineswegs aber wahrscheinlichkeits-theoretisch. Die Aufstellung von Beweiskriterien (bzw. Entscheidungskriterien) gehört nicht in den Aufgabenbereich der Wahrscheinlichkeitstheorie (denn diese liefert lediglich den konsistenten Übergang von beliebigen Annahmen zu anderen Annahmen), sondern kann nur im Rahmen einer (juristischen) Entscheidungstheorie explizit angegangen werden³⁾

Auf S.71 schreibt Cohen, daß die mathematische Wahrscheinlichkeitstheorie außerdem "zu restriktiv ist", denn danach kann es - entgegen der Rechtspraxis - sein, daß auch dann, wenn alle Stufen eines

1) Cohen bezieht sich bei dieser Forderung auf Wigmore, A Treatise on the Anglo-American System of Evidence in Trials at Common Law, (1940).

2) Eine ähnliche Kritik ist bei Schum, S.481, zu finden.

3) Cohen hält die Unfähigkeit der Wahrscheinlichkeitstheorie, (juristische) Beweiskriterien aufzustellen bzw. zu erklären und die Auffassung, daß dies sowieso nicht die Aufgabe der Wahrscheinlichkeitstheorie, sondern eine juristische Entscheidungsfrage ist, für ein Mysterium (S.71): "They have to claim that the non-transitivity of proof on the balance of probability is a purely legal matter...But such a treatment of the difficulty is unsatisfactory because it creates a mystery which it does not resolve, viz. what precisely is the explanation of the restriction?"

mehrstufigen Beweises, die der letzten vorangehen, mit "an Sicherheit grenzender Wahrscheinlichkeit" nachgewiesen sind, die Endwahrscheinlichkeit trotzdem sehr gering ist; z.B. kann in einem zweistufigen Beweis $P(R/Q) = 1$ und $P(S/R)$ auch sehr hoch, dagegen $P(S/Q) = 0$ sein. Als Beispiel mag der von Cohen angeführte Fall (S.71) dienen:

Die Wahrscheinlichkeit, daß eine Person ein Einwohner Englands ist unter der Bedingung, daß sie ein Einwohner von Oxfordshire ist, ist 1, und die Wahrscheinlichkeit dafür, daß eine Person nicht Einwohner von Oxfordshire ist unter der Bedingung, daß sie in England wohnt, ist auch ziemlich hoch. Die Wahrscheinlichkeit, daß die Person außerhalb Oxfordshire wohnt unter der Bedingung, daß sie Einwohner von Oxfordshire ist, ist Null.

Es ist fraglich, ob Fälle, die diese mathematische Struktur (d.h. $P(R/Q) = 1$, $P(S/R)$ (sehr) hoch, aber $P(S/Q) = 0$) haben, für die Rechtspraxis in der Tat relevant sind. Sie sind auch ziemlich schwer zu finden. Man kann allerdings in Analogie zu dem von Cohen angeführten Beispiel ein ähnliches "quasi-juristisches" konstruieren:

Die Wahrscheinlichkeit, daß der Angeklagte der Vater ist, unter der Voraussetzung, er hat die gleichen Blutmerkmale wie das Kind und das Kind hat nicht die gleichen Blutmerkmale wie seine Mutter, ist 1 (wir nehmen an, daß der Angeklagte der einzige Mann ist, der in Frage kommt); die Wahrscheinlichkeit, daß das Kind nicht die gleichen Blutmerkmale wie der Angeklagte aufweist, falls der Angeklagte der Vater ist, ist 0.5 (denn es kann ebenso gut auch Blutmerkmale seiner Mutter erben; im Unterschied zur ersten Wahrscheinlichkeitsaussage, geht hier nicht die Voraussetzung ein, daß das Kind nicht die Blutmerkmale seiner Mutter hat); die Wahrscheinlichkeit, daß das Kind nicht die gleichen Blutmerkmale hat wie der Angeklagte unter der Voraussetzung, daß der Angeklagte die gleichen Blutmerkmale und die Mutter nicht die gleichen Blutmerkmale wie das Kind aufweisen, ist Null.

Die Richtigkeit eines solchen Schlusses wird kaum jemand (und auch nicht ein Jurist) bestreiten. An derartig 'unergiebigem' Schlüssen hat die Rechtsprechung aber gewiß kein Interesse.

Für die wahrscheinlichkeitstheoretische Beurteilung einer empirischen Hypothese aufgrund aller zur Verfügung stehenden relevanten Daten ist der Schluß von $P(R/Q)$ und $P(S/R)$ auf $P(S/Q)$ nicht korrekt¹⁾. Denn in einem solchen Schluß werden eben nicht alle (wahrscheinlichkeitstheoretisch) relevanten Informationen berücksichtigt. Will man S aufgrund von Q unter Berücksichtigung der unsicheren Information R beurteilen, so muß (wiederum) die B_{erw} -Formel angewandt

1) Denn dieser Schluß gilt nicht allgemein, sondern nur unter bestimmten (von Cohen allerdings nicht erwähnten) sehr starken Voraussetzungen (vgl. die nächste Seite).

werden, die hier wie folgt lautet:

$$P(S/Q) = P(S/R \wedge Q)P(R/Q) + P(S/\neg R \wedge Q)P(\neg R/Q).$$

So gesehen ist der Schluß in Cohens Beispiel (bzw. im oben angeführten analogen) korrekterweise auch auf diese Weise rekonstruierbar.

Da in Cohens Beispiel $S = \neg Q$ ist, erhalten wir:

$$P(S/R \wedge Q) = P(\neg Q/R \wedge Q) = 0; \quad P(S/\neg R \wedge Q) = P(\neg Q/\neg R \wedge Q) = 0;$$

außerdem gilt: $P(\neg R/Q) = 0$.

Somit folgt:

$$P(S/Q) = P(\neg Q/Q) = 0 \cdot 1 + 0 \cdot 0 = 0.$$

Der Schluß von $P(R/Q)$ und $P(S/R)$ auf $P(S/Q)$ gilt nur, wenn $R \subseteq Q$ und $S \subseteq R$, d.h. wenn sowohl R Teilmenge von Q wie auch S Teilmenge von R ist, denn dann können wir folgendes ableiten:

$$P(S/R \wedge Q) = P(S/R); \quad P(S/\neg R \wedge Q) = 0;$$

somit gilt insgesamt:

$$P(S/Q) = P(S/R)P(R/Q) + 0 \cdot P(\neg R/Q) = P(S/R)P(R/Q).$$

Ist $P(R/Q) = 1$ und $P(S/R)$ "sehr hoch", so kann $P(S/Q)$ niemals 0 sein, denn es folgt durch Einsetzen in die obige Formel: $P(S/Q) = P(S/R) =$ "sehr hoch". D.h. für die von Cohen kritisierte Schlußfolgerung kann gar nicht diese Formel gelten. Hingegen konnten wir mit der B_{erw} -Formel das im Cohenschen Sinne unerwünschte Ergebnis ableiten, allerdings unter der Voraussetzung, daß $S = \neg Q$ ist¹⁾

Kurz, die von Cohen kritisierten Fälle sind trivial bzw. uninteressant.

1) Mit der Voraussetzung: $P(R/Q) = 1$ gilt nur dann $P(S/Q) = P(S/R \wedge Q)P(R/Q) + P(S/\neg R \wedge Q)P(\neg R/Q) = 0$, wenn $P(S/R \wedge Q) = 0$ ist. D.h. auch dann, wenn $P(S/R) \neq 0$ ist, kann die Hinzunahme eines neuen Datums bewirken, daß das Ereignis aufgrund des alten und des neuen Datums logisch unmöglich bzw. praktisch ausgeschlossen wird. Es ist schwer zu sehen, wieso solche selbstverständlichen Schlüsse von Juristen nicht akzeptiert werden sollen.

Zusammenfassung

(Wir wollen hier noch einmal die Argumente Cohens und die Gegenargumente kurz wiedergeben).

Im Unterschied zu der ersten von Cohen aufgestellten Paradoxie, die gegen die Struktur des Wahrscheinlichkeitsbegriffs (genauer: gegen das wahrscheinlichkeitstheoretische Multiplikationsprinzip) gerichtet ist, werden hier die Anwendungsmöglichkeiten der Wahrscheinlichkeitstheorie kritisiert. So wirft Cohen der Wahrscheinlichkeitstheorie vor, daß sie für die Anwendung in der Rechtspraxis - wenn es sich um einen fortgesetzten (bzw. mehrstufigen) Beweis handelt - einerseits zu liberal und andererseits zu restriktiv ist.

Liberal ist sie, weil sie auch dann den Beweis der Sachverhaltshypothese mit überwiegender Wahrscheinlichkeit erlaubt, wenn nicht alle Stufen eines solchen Beweises, die der letzten vorangehen mit der (vor Gericht geforderten) "an Sicherheit grenzenden Wahrscheinlichkeit" nachgewiesen sind.

Restriktiv ist sie, weil sie zur Nicht-Ableitung der Sachverhaltshypothese - bzw. eines genügend hohen Wahrscheinlichkeitsgrades für sie - auch dann führen kann, wenn jede Stufe des Beweises, die der letzten vorangeht, mit Sicherheit bzw. mit sehr hoher Wahrscheinlichkeit festgestellt ist.

Die Kritik, daß die Wahrscheinlichkeitstheorie zu liberal für die Rechtspraxis sein soll, wurde ihrerseits kritisiert. Es sind v.a. zwei Gründe, die die Einwände Cohens zurückweisen lassen:

1. Cohen gibt kein Explikat des für seine Kritik so wichtigen Begriffs "mit an Sicherheit grenzender Wahrscheinlichkeit" an. Wir haben aber gesehen, daß je nachdem, wie man diesen Begriff interpretiert, er auch verschiedene Konsequenzen für die Rechtspraxis hat.

Falls man ihn als ein Beweiskriterium auffaßt, daß für ein bestimmtes Individuum in einer konkreten Situation eine Wahrscheinlichkeit ausdrückt, die für die Verurteilung (bzw. eine weitere Schlußfolgerung, einen weiteren Beweis) ausreicht, so ist nicht einzusehen, warum die Anwendung der Wahrscheinlichkeitstheorie mit der Rechtspraxis nicht konsistent sein soll.

Faßt man dagegen diesen Begriff in dem Sinne auf, daß "nur erkenntnistheoretische Zweifel bestehen", so hat das für die Rechtspraxis die Konsequenz, daß alle Beweise, die auf unsicheren Fakten (Beweismitteln) beruhen - v.a. Zeugenbeweise - von vornherein zugunsten des Angeklagten entschieden werden müssen. Da dies in der Rechtspraxis gewöhnlich nicht der Fall ist, bedeutet das, daß man sich entweder in der Praxis an diese Forderung nicht hält, oder die "mit an Sicherheit grenzende Wahrscheinlichkeit" anders auffaßt als Cohen (bzw. die theoretisierende Rechtsprechung, was empirische Untersuchungen bestätigen, vgl. S.150ff dieser Arbeit).

2. Der Vorwurf Cohens, die Wahrscheinlichkeitstheorie formuliere und begründe nicht die in der Rechtspraxis geläufigen Beweiskriterien, ist zurückzuweisen. Denn die Aufstellung und Begründung von Beweis- bzw. Entscheidungskriterien gehört in einen anderen Kontext - etwa den einer juristischen Entscheidungstheorie -, keineswegs aber in den der Wahrscheinlichkeitstheorie, deren Aufgabe es ist, lediglich einen Rahmen zu geben, innerhalb dessen man von bestimmten Annahmen zu anderen Annahmen auf konsistente Weise übergehen kann.

Auch die Kritik Cohens, daß die Wahrscheinlichkeitstheorie zu restriktiv sei, ist nicht gerechtfertigt. Sie konfrontiert darüber hinaus die Auffassung Cohens von der "Liberalität der Rechtspraxis" mit Begründungsschwierigkeiten.

Denn man muß überhaupt erst zeigen, daß die von Cohen angesprochenen - wie er schreibt - 'restriktiven' mathematischen (von uns jedoch als trivial erkannten) Schlüsse in der Rechtspraxis eine Rolle spielen. Cohen zeigt das nicht¹⁾ Wir haben gesehen, daß man in Analogie zum Cohenschen Beispiel (vgl. S.185 dieser Arbeit) auch 'quasi-juristische' Beispiele konstruieren kann, an deren Richtigkeit niemand (auch nicht ein Jurist) zweifeln wird (S.185). Solche Fälle sind jedoch für die Rechtspraxis ohne Bedeutung.

Aber auch dann, wenn man in der Rechtspraxis mehrstufige (nicht-triviale) Beweise finden würde, die die von Cohen problematisierte Schlußstruktur haben, und die aufgrund der Wahrscheinlichkeitstheorie zu nicht mit der juristischen Intuition konsistenten (weil für die Rechtspraxis zu restriktiven) Ergebnissen führen würden, müßte man aus juristischer Sicht begründen, warum man ähnliche Schlüsse in anderen Bereichen des induktiven Rasonierens als richtig akzeptiert (vgl. das Beispiel von Cohen und dasjenige von mir S.185 dieser Arbeit), die Restriktionen, die aufgrund solcher Schlüsse entstehen, für die Rechtspraxis aber für falsch hält.

Eine juristische Intuition zu rechtfertigen, die in anderen induktiven Bereichen sowohl zu contra-intuitiven als auch zu nicht rationalen Ergebnissen führen würde, ist jedoch schwierig.

1) Cohens Standpunkt ist normativ: Er sagt zwar, daß bestimmte juristische Beweise diese oder jene Struktur haben müssen, verzichtet aber darauf, dies rational zu rechtfertigen bzw. durch empirische Beispiele zu belegen.

3.Paradoxie: die Negation¹⁾

Nach dem Negationsprinzip ist die Wahrscheinlichkeit für die Nicht-täterschaft des Angeklagten gleich $1-x$, wenn die Wahrscheinlichkeit für seine Täterschaft gleich x ist. Für das (zivilrechtliche) Beweiskriterium "mit überwiegender Wahrscheinlichkeit" bedeutet das, daß $x > 1-x$ sein muß, will man den Angeklagten verurteilen. x kann somit jeder Wahrscheinlichkeitswert > 0.5 sein, wie etwa 0.7 oder aber auch 0.501; dies impliziert nach Cohen aber eine Ungerechtigkeit, vgl.S.75:

"Suppose the threshold of proof in civil cases were judicially interpreted as being at the level of a mathematical probability of 0.501. Would not judges thereby imply acceptance of a system in which the mathematical probability that the unsuccessful litigant deserved to succeed might sometimes be as high as .499? This hardly seems the right spirit on which to administer justice".

Zur Illustration führt Cohen folgendes Beispiel an (S.74):

Angenommen, bei einer Rodeoveranstaltung haben 499 Personen den Eintritt bezahlt, 1000 jedoch wurden auf den Sitzen gezählt. A sei eine davon. Weiter sei angenommen, daß keine Eintrittskarten verteilt wurden, so daß kein Beweisstück vorliegt, ob A den Eintritt bezahlt oder nicht bezahlt hat; so daß die Wahrscheinlichkeit aufgrund der vorhandenen Information dafür, daß A den Eintritt nicht bezahlt hat, 0.501 beträgt. In diesem Fall kann somit die Veranstaltungsadministration gegen A klagen und gewinnen, weil die Wahrscheinlichkeit zu ihren Gunsten überwiegt?²⁾

Die Einwände, die Cohen gegen die Wahrscheinlichkeitstheorie erhebt, sind somit folgende:

1. Die wahrscheinlichkeitstheoretische Interpretation des Beweiskriteriums "mit überwiegender Wahrscheinlichkeit" führt aufgrund des Negationsprinzips dazu, daß eine ziemlich hohe Wahrscheinlichkeit für die Unschuld des Angeklagten sprechen kann, obwohl

1) Cohens Bezeichnung: "The Difficulty about Negation", S.74.

2) Vgl. *ibid.*, S.75: "The mathematicist theory would apparently imply that in such circumstances the rodeo organizers are entitled to judgment against A for the admission-money, since the balance of probability (and also the difference between prior and posterior probabilities) would lie in their favour". Und weiter: "Indeed, if the organizers were really entitled to judgment against A, they would presumably be equally entitled to judgement against each person in the same situation as A. So they might conceivably be entitled to recover 1,000 admission-moneys, when it was admitted that 499 had actually been paid. The absurd injustice of this suffices to show that there is something wrong somewhere. But where?"

man sich für seine Schuld entscheidet (etwa 0.499, wenn die Wahrscheinlichkeit für seine Schuld 0.501 beträgt und man diesen Wert als Beweiskriterium "mit überwiegender Wahrscheinlichkeit" akzeptiert).

2. Die Wahrscheinlichkeitstheorie führt dazu, daß jede Person an der Stelle von A unter den gleichen Umständen verurteilt wird, weil die relevanten Informationen, die hier vorliegen, statistischer Art sind.

(1) Als Antwort auf den ersten Einwand bietet sich gleich der Vorschlag an, daß ein Individuum - falls ihm eine bestimmte Wahrscheinlichkeit (etwa 0.501) in einer konkreten Situation nicht ausreichend für die Verurteilung erscheint - den Wahrscheinlichkeitswert seines Beweiskriteriums erhöhen oder die Beweisgrundlage verstärken soll (beträgt die Wahrscheinlichkeit für die Schuld des Angeklagten aufgrund der Beweisgrundlage z.B. 0.9, so spielt es dann keine Rolle, daß die Grenze des Beweiskriteriums für das Individuum für diesen Fall etwa bereits bei 0.5 liegt).

Cohen lehnt den ersten Vorschlag ab. Die Gründe dafür sind:

(i) Die Erhöhung des Beweiskriteriums verstärke die Paradoxie der Konjunktion. Denn dann dürften die Wahrscheinlichkeiten für zwei einfache Sachverhaltshypothesen z.B. nicht unter 0.8 liegen, will man als Wahrscheinlichkeit für die zusammengesetzte Sachverhaltshypothese etwa ≈ 0.7 (als Beweiskriterium) haben.

(Zur Lösung der Paradoxie der Konjunktion vgl. S. 172ff dieser Arbeit; dieser Einwand kann somit nicht akzeptiert werden).

(ii) Die Erhöhung des Beweiskriteriums würde die Grenze zwischen den beiden Beweiskriterien "mit überwiegender Wahrscheinlichkeit" und "mit an Sicherheit grenzender Wahrscheinlichkeit" verwischen.

Auch dieser Einwand ist nicht stichhaltig angesichts der Tatsache, daß es kein objektives Kriterium gibt, das die beiden obigen Ausdrücke zahlenmäßig präzisieren kann; welcher Wahrscheinlichkeitswert in einem konkreten Fall für die Verurteilung ausreicht, entscheidet das Individuum ganz allein aufgrund seiner subjektiven Einschätzung der Situation. Und wie bereits gezeigt wurde¹⁾, ist der Unterschied zwischen den beiden Beweiskriterien in der Tat

1) Vgl. § 2 dieses Kapitels.

gering (die beiden variieren im Intervall von 0.5 bis 1; vgl. die Untersuchungen vom Simon&Mahan, Eggleston, Nagel&Neef). Es ist sogar der Fall ohne weiteres denkbar, in dem für ein Individuum ein Wahrscheinlichkeitswert etwa von 0.7 die "mit an Sicherheit grenzende Wahrscheinlichkeit" darstellt, für ein anderes Individuum dagegen (im gleichen oder in einem anderen Fall) nicht einmal die Wahrscheinlichkeit von 0.9 eine "überwiegende Wahrscheinlichkeit" bedeutet¹⁾

(iii) Aber auch dann, wenn man den Beweiswert für die "überwiegende Wahrscheinlichkeit" erhöht (wenn dieser Wert etwa 0.8 ist), so bliebe genug Raum für den Zweifel, was nach Cohen "skandalös"²⁾ wäre.

Auch diesen Einwand kann man kaum akzeptieren, da das Beweiskriterium "mit überwiegender Wahrscheinlichkeit" gerade das verlangt, was Cohen kritisiert, nämlich, daß ein Sachverhalt 'nur' wahrscheinlicher sein soll als der andere, und 0.8 ist in der Tat größer als 0.2.³⁾

(2) Zum zweiten Kritikpunkt kann man lediglich vorbringen, daß er eigentlich nicht so sehr gegen die Wahrscheinlichkeitstheorie an sich gerichtet ist, als vielmehr gegen eine bestimmte Interpretation derselben, nämlich der als Häufigkeits- (bzw. statistischer) Theorie.

Interpretiert man den juristischen Wahrscheinlichkeitsbegriff als statistischen Wahrscheinlichkeitsbegriff, so ist nicht einzusehen, warum wir nicht die statistische Information (z.B. im obigen Beispiel), die zwar über das Gesamtkollektiv, nichts jedoch konkret über eine einzelne Person aussagt, unserer Beurteilung des juristischen Falls zugrunde legen sollen. In Wirklichkeit wird jedoch

1) Vgl. Simon&Mahan, Tabellen 4 und 7.

2) Vgl. Cohen, S.76.

3) Vgl. die Kritik von Kaye an dieser Paradoxie (S.38): "One could simply dispute the paradoxical quality of the conundrums Cohen creates. Where, for instance, Cohen insists that to refuse recovery to a civil litigant who had proved that the probability the defendant is liable is .499 "hardly seems the right spirit in which to administer justice", one might reasonably suggest that such an outcome is perfectly appropriate in the uncertain and imperfect world of litigation. There is, in addition, a different and perhaps more decisive rejoinder to Cohen's ambitious argument. Whatever puzzlement is produced by Cohen's "anomalies" is, I think, attributable to an underlying misconception about the role of probability theory in legal proof. Conventional probability theory does not purport to describe the structure of current legal doctrine. Rather, it prescribes how probabilistic evidence should be handled if one is to make the fewest mistakes in predicting the outcome of uncertain events". (Hervorhebung von mir).

kaum ein Richter bereit sein, die Verurteilung nur aufgrund dieser Information auszusprechen¹⁾

Die Situation ändert sich, wenn wir den juristischen Wahrscheinlichkeitsbegriff subjektivistisch interpretieren. Denn das subjektive Wahrscheinlichkeitsurteil ist nichts anderes als die quantitative Präzisierung des qualitativen (intuitiven) Urteils des Individuums²⁾ (Hält das Individuum die vorliegenden Informationen für die Beurteilung eines konkreten Falls für irrelevant - wie es wohl viele im obigen Beispiel tun würden - so braucht es sie auch aufgrund des subjektiven Wahrscheinlichkeitsbegriffs nicht zu berücksichtigen)³⁾

Zusammenfassung

Cohen dritte Paradoxie ist anscheinend nicht so sehr gegen die Wahrscheinlichkeitstheorie als gegen die statistische Interpretation des Wahrscheinlichkeitsbegriffs gerichtet (auch wenn Cohen das nicht zugeben würde).

Interpretiert man den juristischen Wahrscheinlichkeitsbegriff statistisch, so besteht kein Grund, wieso wie den Angeklagten nicht aufgrund der statistischen Informationen verurteilen sollen, wenn ein bestimmter Beweiswert erreicht ist. (So ergibt sich im von Cohen angeführten Beispiel die paradoxe Situation, daß der Angeklagte aufgrund einer einzigen Information verurteilt wird, die mit ihm konkret nichts zu tun hat und die seine Schuld nur zu 0.501 wahrscheinlich macht; kaum jemand würde den Angeklagten unter solchen Umständen verurteilen).

Aus subjektivistischer Sicht ergibt sich dagegen keine Schwierigkeit, denn dann liegt es einzig und allein am Individuum, zu entscheiden, ob die vorliegenden Informationen für die Wahrscheinlichkeits-

1) Das meint übrigens auch Cohen auf S.79. Es sei hier daran erinnert, daß bereits Finkelstein & Fairley (die eigentlich den statistischen Standpunkt vertreten), bei ihrer Anwendung des Bayes-Ansatzes bei der Bestimmung der Identität, darauf hingewiesen haben, daß es unzulässig ist, die Hypothese allein aufgrund statistischer Daten zu beurteilen; vgl. S.123 dieser Arbeit.

2) Vgl. 1.Kap., §2 dieser Arbeit.

3) Es ist interessant festzustellen, daß Cohen die subjektivistische Interpretation des Wahrscheinlichkeitsbegriffs in diesem Zusammenhang überhaupt nicht ernsthaft diskutiert. Vgl. seine, bei den Subjektivisten auf Ablehnung stoßende Behauptung (S.77): "Even if you think of the .8 threshold subjectively, as a degree of belief, you presumably still have to adjust your degree of belief to the objective frequencies".

theoretische Beurteilung des konkreten Einzelfalls relevant sind oder nicht (bzw. ob sie ausreichen oder nicht).

Was die Beweiskriterien betrifft, so muß das Individuum selbst entscheiden, welcher Wahrscheinlichkeitswert im konkreten Fall für die Verurteilung ausreicht.

Akzeptiert das Individuum als solchen Wahrscheinlichkeitswert etwa den von 0.501 (wozu es wohl gemerkt nicht gezwungen ist) und erreicht die Sachverhalts-hypothese eine Wahrscheinlichkeit von mindestens 0.501, so muß es sich dementsprechend verhalten (und u.a. berücksichtigen, daß eine 'ziemlich hohe' und, wie Cohen sagen würde, "skandalöse" Wahrscheinlichkeit von 0.499 für die Negation der Hypothese vorliegt). Die Wahrscheinlichkeitsverteilung von 0.501 für die Hypothese und 0.499 dagegen, kann man übrigens auch so interpretieren, daß hier eine (eigentlich nicht so seltene) Situation eingetreten ist, wo das Individuum sich kaum zwischen beiden Hypothesen entscheiden kann und wo es letztlich doch eine davon wählt.

(Wir werden im zweiten Teil dieser Arbeit die Wahl der strafrechtlichen Handlungen entscheidungstheoretisch in dem Sinne präzisieren, daß wir nicht nur (wie es oben der Fall ist) für die Entscheidungswahl den Wahrscheinlichkeitswert für die Sachverhaltshypothese berücksichtigen, sondern auch, wie wünschenswert die Konsequenzen solcher Entscheidungen sind).

4. Paradoxie: Beweis "mit an Sicherheit grenzender Wahrscheinlichkeit"¹⁾

Diese Paradoxie besteht nach Cohen darin, daß die Wahrscheinlichkeitstheorie den Ausdruck "mit an Sicherheit grenzender Wahrscheinlichkeit" nicht präzisieren kann und außerdem immer zum Anzweifeln einer Schlußfolgerung führt, falls diese nicht mit Sicherheit festgestellt ist. Das sei für die Rechtspraxis unannehmbar. Mit Cohens Worten (S.82f):

"We are more inclined to hold that a particular conclusion falls short of certainty because there is a particular, specifiable reason for doubting it, than to hold that it is reasonable to doubt the conclusion because it falls short of certainty. Even if a scala of mathematical probability were to be used for assessing how close a particular conclusion was to certainty, the crucial reason for doubting the conclusion would very often not be the fact that the probability was no higher than, say, .95, but rather the specific item in, or feature of, the evidence that prevented this probability from rising any higher. What makes it reasonable to doubt the guilt the accused, perhaps, is the shifty demeanour of one of the key witnesses against him or the absence of any evidence of motive. Correspondingly, even if a scala of mathematical probability could be used for assessing how close a particular conclusion was to certainty, its assessments would very often be quite superfluous as reasons for doubting guilt. The proof's degree of validity could be judged without regard to the scala of mathematical probability. The scala would be otiose. This is another price that has to be paid by those who insist on conceiving all judicial probabilities as mathematical ones".

Wenn das alles ist, was gegen die Wahrscheinlichkeitstheorie angeführt werden kann - und mehr schreibt Cohen in der Tat nicht - so ist wiederum schwer einzusehen, was an ihr paradox sein soll. Denn was (wie bereits mehrmals erwähnt) ein Individuum in einem konkreten juristischen Fall für eine für die Verurteilung ausreichende Wahrscheinlichkeit hält (aber auch, ob der Sachverhalt diese Wahrscheinlichkeit bzw. welchen Beweiswert dieser erreicht hat), wird von ihm sowohl im qualitativen (herkömmlichen) wie auch im quantitativen Fall subjektiv entschieden. (So gesehen, ist auch der Wahrscheinlichkeitswert von 0.95 (im obigen Zitat) nichts anderes als die Präzisierung des intuitiven Urteils des Individuums - eben, weil ein Grund vorliegt (wie etwa verdächtiges Benehmen des Zeugen oder das Fehlen des Motivs usw.), die Schlußfolgerung anzuzweifeln). Diese Wahrscheinlichkeit ist also nichts weiter als ein Ausdruck dafür, daß es "a particular, specifiable reason" für den

¹⁾ Cohens Bezeichnung: "The Difficulty about Proof beyond Reasonable Doubt", S.82.

Zweifel an der Schlußfolgerung gibt (diese Wahrscheinlichkeit braucht übrigens nicht - wie mehrere empirische Untersuchungen entgegen der Auffassung Cohens bestätigen¹⁾ - nur "short of certainty" zu sein).

Außerdem führt - wie die Untersuchung von Simon&Mahan gezeigt hat²⁾ - die Quantifizierung der intuitiven Urteile zu einer vorsichtigeren Entscheidungspraxis. Darüber hinaus ist die Zuordnung von Wahrscheinlichkeitswerten zu Urteilen unumgänglich, will man juristische Fälle entscheidungstheoretisch lösen, d.h. unter der Berücksichtigung nicht nur der Unsicherheit bzgl. des Sachverhalts, sondern auch dessen, wie wünschenswert bzw. schwerwiegend Konsequenzen u.a. aus Fehl-Entscheidungen sind³⁾

Die Zuordnung von Wahrscheinlichkeitswerten zu intuitiven Urteilen ist somit aus folgenden Gründen von Nutzen⁴⁾

1) Sie gewährleistet die Präzisierung des qualitativen (intuitiven) Urteils (bzw. "Gefühls"). (So gesehen ist auch das Wahrscheinlichkeitsurteil "mit an Sicherheit grenzender Wahrscheinlichkeit" eine Präzisierung des entsprechenden qualitativen subjektiven Urteils).

2) Sie ermöglicht eine bessere intersubjektive Nachvollziehbarkeit der einzelnen Urteilsbildung.

3) Sie ermöglicht außerdem eine entscheidungstheoretische (d.h. rationale) Lösung des juristischen Falls. (Es wird dabei nicht nur die Unsicherheit bei der Feststellung des Sachverhalts berücksich-

1) Vgl. §2 dieses Kapitels.

2) Vgl. S.150ff dieser Arbeit.

3) Der Unterschied zwischen der Entscheidung nach dem Sachverhalt (falls seine Wahrscheinlichkeit für die Verurteilung ausreicht) und der entscheidungstheoretischen Lösung sei hier kurz erläutert: Reicht im ersten Fall z.B. eine Wahrscheinlichkeit von 0.8 für den Sachverhalt für die Verurteilung und erreicht der Sachverhalt diese Wahrscheinlichkeit, so entscheidet man auch gleich für eine Verurteilung. Im zweiten Fall dagegen werden darüber hinaus auch die Konsequenzen verschiedener Entscheidungen beurteilt. Wiegt z.B. die Konsequenz einer nichtkorrekten Verurteilung schwerer als sonstige Konsequenzen (bzw. ist sie sehr unerwünscht), so kann es passieren, daß der Freispruch gewählt wird, obwohl der Sachverhalt eine für die Verurteilung ausreichende Wahrscheinlichkeit erreicht hat. Näheres darüber im II. Teil dieser Arbeit.

4) Für Cohen ist eine solche Zuordnung selbstverständlich "überflüssig" bzw. "zwecklos"; vgl. das o. Zitat.

tigt, sondern auch, wie schwerwiegend (wünschenswert) die Konsequenzen der jeweiligen Entscheidungen sind).

4) Sie erlaubt erst einen präziseren Vergleich zwischen dem für einen konkreten Fall angenommenen Beweiswert (d.h. einer für die Verurteilung ausreichenden Wahrscheinlichkeit) und dem vom Sachverhalt erreichten Wahrscheinlichkeitswert.

In diesem Zusammenhang ist es interessant, auf die Rolle einzugehen, die die Wahrscheinlichkeitstheorie nach Cohens Meinung in der Rechtspraxis spielen kann. Cohen lehnt die mathematische Wahrscheinlichkeitstheorie - wobei er darunter ausschließlich die statistische Theorie versteht (S.82: "Not that a high statistical probability is necessarily useless") - nicht völlig ab; so können nach Cohen statistische Beweise einen Teil der Beweisgrundlage bilden, auf der die Schlußfolgerung beruht, nicht jedoch auch Wahrscheinlichkeitswerte für die Schlußfolgerung selbst liefern¹⁾

Zur Illustration gibt Cohen das folgende Beispiel an (S.83f):

Angenommen, A sei wegen Mordes an B angeklagt. Weiter sei angenommen, daß ein Teil der Beweismittel gegen A ein Drohbrief ist, der im Haus von B gefunden wurde und der - obwohl nicht unterschrieben - auf einer Schreibmaschine geschrieben wurde, die die gleichen Merkmale aufweist wie die von A²⁾. Es sei angenommen, daß Experten drei solche (wechselseitig unabhängigen) Merkmale sowohl an der Schreibmaschine, auf der der Brief geschrieben ist wie auch an der Schreibmaschine von A annehmen. Jedes dieser Merkmale komme nur einmal in 10 Schreibmaschinen vor. Daraus folgt, daß die Wahrscheinlichkeit dafür, daß eine Schreibmaschine diese drei Merkmale besitzt, 0.001 ist; und die Wahrscheinlichkeit, daß ein zufällig gewähltes Paar von Schreibmaschinen, unabhängig voneinander, diese drei Merkmale besitzen wird, ist 1/1000000.

Cohen zeigt uns also, daß die Wahrscheinlichkeit für das gemeinsame Vorkommen der drei in Frage kommenden Merkmale bei einer zufällig gewählten Schreibmaschine 1/1000 und bei zwei zufällig gewählten (voneinander unabhängigen) Schreibmaschinen 1/1000000 ist. Es

1) Was übrigens auch der in dieser Arbeit vertretenen Auffassung entspricht: die statistischen Wahrscheinlichkeiten können lediglich als ein Teil der Beweisgrundlage dienen. Es wird aber von uns im Unterschied zu Cohen angenommen, daß die Relation zwischen der Beweisgrundlage und der Schlußfolgerung (d.h. der Sachverhaltshypothese) mittels des subjektiven Wahrscheinlichkeitsbegriffs zu expizieren ist.

2) Vgl. Cohen S.83f: "...a threatening letter...which, though unsigned, has been typewritten on a machine that has the same peculiarities... as A's typewriter". (Nebenbei bemerkt: Was haben diese beiden Tatsachen miteinander zu tun? Hervorhebung von mir).

ist jedoch unklar, wen Cohen mit der folgenden Stelle meint (S.84):

"It is tempting thence to infer a .999,999 mathematical probability that the letter's having been typed on a machine with those peculiarities is dependent on the fact that A has such a machine, and therefore apparantly a .999,999 mathematical probability that A typed the letter".

Gegen die Proponenten der Wahrscheinlichkeitstheorie kann dieser Angriff kaum gerichtet sein. Denn diese haben derartige Fehler - d.h. die Identifizierung der Wahrscheinlichkeit dafür, daß ein zufällig gewähltes Individuum (oder Objekt) die in Frage kommenden Merkmale nicht besitzt (bzw. wie in Cohens Beispiel die Wahrscheinlichkeit, daß wenigstens eine der zwei zufällig und unabhängig voneinander gewählten Schreibmaschinen diese Merkmale nicht besitzt¹⁾ mit der Wahrscheinlichkeit für die Identität des gefundenen Individuums (oder Objekts) mit dem gesuchten (bzw. darüber hinaus in Cohens Beispiel mit der Wahrscheinlichkeit, daß A den Brief geschrieben hat) - bereits zur Genüge ausdiskutiert²⁾

Somit bleiben nur die Juristen.

Für die Berechnung der Identitätswahrscheinlichkeit (d.h. der Wahrscheinlichkeit dafür, daß das gefundene Individuum auch die fraglichen Merkmale am Tatort hinterlassen hat (bzw. die fraglichen Merkmale vom Tatort besitzt) bzw. der Duplizierungswahrscheinlichkeit (d.h. daß es mehr als ein solches Individuum gibt), ist lediglich die Wahrscheinlichkeit von 0.001 (für das Vorkommen der drei (voneinander unabhängigen) Merkmale bei einem zufällig gewählten Individuum (oder Objekt) als Parameter wichtig; die Wahrscheinlichkeit von 1/1000000 dafür, daß zwei zufällig voneinander unabhängig gewählte Schreibmaschinen die Merkmale besitzen, ist völlig überflüssig. Vgl. dazu auch die Kritik von Birmingham³⁾ der in Anspielung auf die oben zitier-

1) Wenn $p = 1/1000000$ die Wahrscheinlichkeit dafür ist, daß bei einem zufällig gewählten (voneinander unabhängigen) Schreibmaschinenpaar sowohl die eine (M_1) wie auch die andere Schreibmaschine (M_2) alle in Frage kommenden Merkmale aufweist, dann ist $1-p = 1-P(M_1 \wedge M_2) = P(M_1 \wedge \neg M_2) \vee P(\neg M_1 \wedge M_2) \vee P(\neg M_1 \wedge \neg M_2) = 999999/1000000$ die Wahrscheinlichkeit dafür, daß wenigstens eine aus dem zufällig gewählten Schreibmaschinenpaar diese Merkmale nicht besitzt. (Man wird wohl kaum - im Gegensatz zur Auffassung Cohens - in Wirklichkeit an der Wahrscheinlichkeit eines solchen Ereignisses Interesse haben. Wofür man sich wirklich interessiert, ist die Wahrscheinlichkeit, daß eine zufällig gewählte Schreibmaschine die in Frage kommenden Merkmale nicht besitzt; diese Wahrscheinlichkeit kann dann als Parameter der von Kingston (S.93ff dieser Arbeit) vorgeschlagenen Binomialformel benutzt werden).

2) Vgl. Kingston, Cullison u.a. und die Diskussion um den Collins-Fall, §1 dieses Kapitels.

3) R.Birmingham, Remarks on 'Probability' in Law: Mostly, A casenote an a Book Review, in: 12 Georgia Law Rev. (1978), S.535-551.

te Stelle "...it is tempting..." schreibt (S.545, Fn.43):

"It is not tempting at all. Consider the airplane passenger who brought a bomb because he was pretty sure it would not go off and reasoned that the probability of one bomb on a plane being p , the probability of two is $p \times p$ ".

Vgl. auch Cohen (S.84):

"What I want to emphasize here is that even if, in the situation described, it can be shown to be highly probable that A typed the letter this high probability is not to be identified with the .999,999 mathematical probability described by the expert witness".

Diese Behauptung ist richtig (wir können sie sogar ergänzen: Auch dann, wenn mit einer sehr hohen Wahrscheinlichkeit festgestellt ist, daß der Brief auf der Schreibmaschine von A geschrieben ist, könnte A trotzdem den Brief nicht geschrieben haben; und weiter: Auch dann, wenn mit einer sehr hohen Wahrscheinlichkeit festgestellt ist, daß A den Brief geschrieben hat, könnte er trotzdem nicht der Mörder sein), allerdings mißverständlich: Denn nach der vehementen und weltbekannten Diskussion um den Collins-Fall ist kaum zu erwarten, daß ein wirklicher Experte die Wahrscheinlichkeit 0.999999 (für das unwichtige Ereignis, daß wenigstens eine aus dem Schreibmaschinenpaar die Merkmale nicht besitzt) jemals als relevant ableiten würde.

Cohen gibt übrigens richtig an, warum der statistische Wahrscheinlichkeitsbegriff¹⁾ für die Ableitung der Schlußfolgerung, daß A den Brief geschrieben hat, ungeeignet ist; denn dieser läßt die für diesen konkreten Fall relevanten Umstände außer Acht²⁾.

Aus diesem Grund die Wahrscheinlichkeitstheorie für die Bestimmung der Sachverhaltsannahme abzulehnen - wie Cohen es tut - ist jedoch voreilig. Denn die subjektiven Wahrscheinlichkeitstheorie löst gerade die von Cohen angesprochenen Schwierigkeiten.

Zusammenfassung

Bei dem Begriff "mit an Sicherheit grenzender Wahrscheinlichkeit" handelt es sich um ein subjektives Beweiskriterium. Es ist daher klar, daß jener nicht mittels des Begriffs der objektiven Wahrscheinlichkeit expliziert werden kann. Cohens Kritik der Wahrscheinlichkeitstheorie richtet sich aber gerade gegen die objektive (statistische) Auffassung. So schreibt er - was nur in bezug auf die statistische Wahrscheinlichkeitstheorie richtig ist - daß die Wahr-

1) Cohen verwendet dafür mißverständlicherweise die Bezeichnung "mathematische Wahrscheinlichkeit"; wie bekannt, kann die mathematische Wahrscheinlichkeit nicht nur statistisch, sondern auch subjektivistisch interpretiert werden.

2) Vgl. Cohen (S.85): "What counts at this stage in the proof is not just the evidence that there is a high mathematical probability of such a connection in general, but also the evidence that the probability is relevant to the particular case".

scheinlichkeitstheorie nicht berücksichtigt, daß es besondere Gründe gibt, an der Gewißheit einer bestimmten Schlußfolgerung (Sachverhaltsannahme) zu zweifeln. Auch andere Äußerungen Cohens bestätigen, daß er bei seiner Kritik eine für die Rechtspraxis nicht adäquate Interpretation des Wahrscheinlichkeitsbegriffs meint bzw. angreift.

Im Rahmen der subjektiven Wahrscheinlichkeitstheorie ist es dagegen nicht schwer, den Begriff "mit an Sicherheit grenzender Wahrscheinlichkeit" zu interpretieren. Denn der Wahrscheinlichkeitswert ist dann nichts anderes als die quantitative Präzisierung des qualitativen (intuitiven) Beweiswertes, der vom Individuum in einem konkreten juristischen Fall als für eine Verurteilung ausreichend angenommen wird. So gesehen, trägt eine subjektive Wahrscheinlichkeit unter 1 der Tatsache Rechnung, daß es irgendwelche (besonderen) Gründe gibt, an der Sicherheit der Schlußfolgerung zu zweifeln¹⁾

Außerdem ist die Behauptung Cohens, die Quantifizierung juristischer Urteile sei (aus dem obigen Grund) überflüssig bzw. zwecklos, zurückzuweisen²⁾

1) Für Cohen war die Annahme des Gegenteils davon gerade der Grund dafür, den Wahrscheinlichkeitsbegriff als Explikat für den Begriff "mit an Sicherheit grenzender Wahrscheinlichkeit" abzulehnen. Vgl. seine ziemlich sophistische Behauptung (S.82): "We are more inclined to hold that a particular conclusion falls short of certainty because there is a particular, specifiable reason for doubting it, than to hold that it is reasonable to doubt the conclusion because it falls short of certainty". Aus subjektivistischer Sicht trägt der Wahrscheinlichkeitswert gerade der Tatsache Rechnung, daß es bestimmte Gründe gibt, die Sachverhaltsannahme anzuzweifeln.

2) Auf S.195f dieser Arbeit haben wir vier Gründe angegeben, wozu die Quantifizierung juristischer Urteile gut sein kann.

5. Paradoxie: das Kriterium¹⁾

Nach Cohen liefert die Wahrscheinlichkeitstheorie kein rationales Kriterium für ihre Anwendung auf die Rechtspraxis. Cohen führt nicht explizit aus, was er unter einem solchen Kriterium versteht bzw. wozu ein solches Kriterium gut sein soll²⁾. Aus dem Kontextzusammenhang ergibt sich, daß er damit offensichtlich ein solches im Sinne der Rechtfertigung einer Interpretation meint.

So lehnt Cohen den statistischen Wahrscheinlichkeitsbegriff ab, weil dieser der Beurteilung juristischer Einzelfälle nicht gerecht werde³⁾ und den logischen Wahrscheinlichkeitsbegriff⁴⁾, weil er wegen der unendlichen Zahl von Bestätigungsfunktionen, keine einheitliche Behandlung der Rechtsfälle garantiere. Die Wahl einer bestimmten Bestätigungsfunktion, die die Anwendung der logischen Wahrscheinlichkeiten auf die Rechtspraxis plausibel machen könnte, könne man aber nicht rechtfertigen⁵⁾.

Auch der Begriff der rationalen Wette als Zuordnungsmethode für Wahrscheinlichkeitswerte wird als für die Rechtspraxis ungeeignet verworfen⁶⁾.

1) Cohen: "The Difficulty about a Criterion", S.87.

2) Vgl. Cohen (S.87): "Jurors and judges must be supposed to have some criterion, or a system of criteria, for the probabilities of proofs, however tacit, imprecise, and intuitively this criterion or system may be. There must be some rational behind their assignments and comparisons of probability - some rational that justifies public confidence, ensures a reasonable degree of agreement, and underwrites the rejection of perverse verdicts. But it is very difficult to say what their criterion can be if the probabilities involved are mathematical ones".

3) *ibid.*, S.88.

4) Zum logischen Wahrscheinlichkeitsbegriff siehe S.50ff dieser Arbeit.

5) Cohen, S.88f.

6) Cohen vermeidet den Ausdruck "subjektive Wahrscheinlichkeit", aber offenbar ist sie mit dieser Kritik gemeint.

Vgl. auch die Kritik von Schum daran (S.477): "Later, in discussion the 'paradox' involving a choice of P_M interpretation for juridical application, he avoids mention of personal or subjective probability as one possible choice. We recall that he does rule out classical, relative frequency, and confirmation-function interpretations. Since the author does not mention the subjective interpretation, presumably he does not rule it out".

An einer früheren Stelle (S.27) nimmt Cohen allerdings (unfaire) Stellung zum subjektivistischen Wahrscheinlichkeitsbegriff: "But, presumably because it is not easy to become so rational, well-informed and conventionally motivated and because, with wins being undiscoverable, betting on open-ended generalizations is hardly an appropriate activity for rational men, few researchers in the natural or social sciences have in fact adopted this personalist approach".

Die Gründe für die Ablehnung des Begriffs der fairen Wette¹⁾ sind folgende:

- 1) Das Schließen von Wetten mit bestimmten Wetteinsätzen hat nur dann Sinn, wenn Ereignisse (Ergebnisse) anders als aufgrund von Daten erfahrbare sind, in einem Rechtsfall können wir aber auf Ereignisse nur aufgrund von relevanten Daten schließen²⁾
- 2) Wetteinsätze müssen spezifiziert werden; das führt dazu, daß die Höhe der Wetteinsätze (bzw. der subjektiven Werte von solchen) die Höhe der Wettquotienten (eines Ereignisses als Maß für dessen Wahrscheinlichkeit) beeinflußt; dies bedeutet für die Rechtspraxis eine Ungerechtigkeit, denn je nachdem, wieviel eingesetzt wird, kann die Entscheidung verschieden ausfallen³⁾

Cohen lehnt also die Wahrscheinlichkeitstheorie ab, weil keine ihrer Interpretationen seiner Meinung nach ihre Anwendung auf die

1) Zum Begriff der (fairen) Wette siehe S.28f dieser Arbeit.

2) Vgl. Cohen (S.90): "...to request a jurymen to envisage a wager on a past event, when, ex hypothesi, he normally already knows all the relevant evidence that is likely to be readily obtainable, is to employ the concept of a wager in a context to which it is hardly appropriate".

Cohen übersieht anscheinend, daß ein stärkerer Einwand gegen den Begriff der Wette nicht der ist, daß Ereignisse (was ihren Wahrheitswert betrifft) anders als aufgrund von relevanten Daten feststellbar sein sollen, sondern daß Ereignisse in juristischen Fällen überhaupt nicht feststellbar sind (d.h. ob solche wahr sind oder falsch). Aus subjektivistischer Sicht würde man allerdings argumentieren, daß es sich hier nicht um tatsächliche, sondern um sog. hypothetische Wetten handelt und daß man bei dem subjektivistischen Wettkonzept nicht an den tatsächlichen Ergebnissen, sondern allein an der Vorgehensweise beim Abschließen von Wetten interessiert ist (dazu noch später auf S.202 Fn.3 dieser Arbeit).

Cohen vertritt freilich eine andere Auffassung, wenn er schreibt, daß man die Wahrheit eines juristischen Ereignisses "sicher" aufgrund der relevanten Informationen erfährt; vgl. S.90: "So there is no sure way of discovering whether the accused is guilty, or the plaintiff's statement of claim is true, except by looking at the relevant evidence" (Hervorhebung von mir).

3) Vgl. Cohen (S.91): "Every jurymen would have to be instructed separately by the judge, in accordance with the judge's estimate of what would be an appropriate sum of money for that jurymen to envisage wagering. Consequently every accused would be at risk not only in relation to the evidence, but also in relation to the judge's estimates of how much importance each jurymen attaches to the gain or loss of this or that sum of money. Such a system certainly does not yet exist anywhere, and its institution seems scarcely likely to promote the ends of justice".

Rechtspraxis rechtfertigt: die objektivistische (statistische) Interpretation, weil sie keine Aussagen über Einzelfälle erlaubt, die logische, weil sie die Wahl einer bestimmten Bestätigungsfunktion nicht rechtfertigen kann; die subjektivistische schließlich, weil der Begriff der Wette bei der Zuordnung von Wahrscheinlichkeitswerten zu Sachverhaltshypothesen in juristischen Fällen nicht anwendbar ist.

Wie wir auf S.28ff dieser Arbeit festgestellt haben, dient der Begriff der Wette dazu, die Axiome des quantitativen Wahrscheinlichkeitsbegriffs zu begründen¹⁾. Dies ist jedoch nicht die einzige Möglichkeit: So können wir die Begründung dieser Axiome auch über die Metrisierung eines komparativen Begriffs der subjektiven Wahrscheinlichkeit vornehmen. Dies ist sogar von größerer Allgemeinheit und größerer systematischer Bedeutung als die Einführung der Wahrscheinlichkeitsaxiome durch die Betrachtung rationaler Wetten. Will man daher die subjektivistische Auffassung kritisieren, so muß man nicht nur zeigen, daß das Wettkonzept, sondern auch das Metrisierungskonzept im Rechtsbereich undurchführbar ist (oder, einfacher: daß die Axiome des qualitativen (komparativen) Wahrscheinlichkeitsbegriffs²⁾ nicht gelten³⁾).

Zugunsten der subjektivistischen Auffassung spricht unbestreitbar, daß sie die von Cohen im Zusammenhang mit der statistischen und logischen Interpretation angesprochenen Schwierigkeiten vermeidet: Denn sie ist auf Einzelfälle anwendbar und sie geht von einer einzigen Wahrscheinlichkeitsfunktion aus.

Die subjektivistische Wahrscheinlichkeitstheorie erfährt außerdem eine weitere Rechtfertigung, nämlich die entscheidungstheoretische: Denn in juristischen Fällen (wie in vielen praktischen) müssen wir unter Unsicherheit entscheiden. Der Grundbegriff der rationalen

1) und sie darüber hinaus intuitiv plausibel zu machen.

2) Vgl. S.24ff dieser Arbeit.

3) Allerdings wird man aus subjektivistischer Sicht auch die Auffassung bestreiten, daß das Konzept der Wetten auf juristische Fälle nicht anwendbar ist (vgl. S.201Fn.2 dieser Arbeit), denn anstatt von Wetten mit realen Einsätzen und realisierbaren Ergebnissen (Ausgängen) kann man immer noch von hypothetischen Wetten ausgehen. (Zum Begriff der hypothetischen Wette siehe S.47, Fn.1 dieser Arbeit). Und schließlich gibt es außer dem Konzept der Wetten auch andere Möglichkeiten (vgl. Anhang), den intuitiven Urteilen Wahrscheinlichkeitswerte zuzuordnen.

Entscheidungstheorie (die Unsicherheit im Sinne von Risiko berücksichtigt) ist - wie bereits erwähnt - aber der subjektive Wahrscheinlichkeitsbegriff.

Zusammenfassung

Cohen vertritt die Auffassung, daß keine Interpretation der Wahrscheinlichkeitstheorie ihre Anwendung in der Rechtspraxis rechtfertigt¹⁾.

Die statistische nicht, weil sie nicht auf juristische Einzelfälle anwendbar ist; die logische nicht, weil sie nicht die Wahl einer bestimmten Bestätigungsfunktion garantiert; und die subjektivistische²⁾ nicht, weil sie von einem in der Rechtspraxis undurchführbaren Wettkonzept ausgeht.

Während Cohens Kritik an den beiden ersten Interpretationen richtig ist, ist diejenige des subjektivistischen Ansatzes zurückzuweisen. U.a. wird die Begründung der wahrscheinlichkeitstheoretischen Axiome über die Metrisierung eines komparativen Begriffs der subjektiven Wahrscheinlichkeit durch Cohens Kritik keineswegs in Frage gestellt.

Die Anwendung der subjektiven Wahrscheinlichkeitstheorie in der Rechtspraxis erfährt ihre stärkste Rechtfertigung, indem man zeigt, daß sie entscheidungstheoretisch relevant ist: Denn in der Rechtspraxis müssen wir nicht nur die in Frage kommenden Sachverhalte subjektiv bewerten, sondern darüber hinaus unter Unsicherheit (Risiko) entscheiden. Der Begriff der rationalen Entscheidungstheorie ist aber der subjektive Wahrscheinlichkeitsbegriff.

1) Wie bereits gesagt, habe ich den Cohenschen vagen Ausdruck "Kriterium" im Sinne eines Rechtfertigungskriteriums verstanden. (Cohen läßt - wie oft - den Leser völlig im unklaren, was er mit seinen Begriffen genau meint).

2) Wie erwähnt (Vgl.S.200 Fn.6 dieser Arbeit), spricht Cohen nicht explizit von der subjektivistischen Wahrscheinlichkeitstheorie, meint aber offensichtlich sie mit seiner Kritik am Wettkonzept.

6. Paradoxie: Korroboration und Konvergenz¹⁾

Nach Cohen ist es sehr schwierig, einen juristischen Beweis wahrscheinlichkeits-theoretisch nachvollziehen, der entweder korrobierende (sich gegenseitig bestärkende) Zeugenaussagen oder konvergierende Indizien enthält. Sehen wir uns dazu die Erklärungen und Beispiele Cohens an²⁾

1. Die Zeugenkorroboration ist immer dann vorhanden, wenn Zeugen unabhängig voneinander zugunsten einer Proposition (bzw. Sachverhaltsannahme) aussagen. So hebt die Tatsache, daß A zugunsten von p aussagt, auch die Wahrscheinlichkeit, daß p der Fall ist und ebenso, wenn B zugunsten von p aussagt³⁾. Zusammen aber erhöhen diese beiden Zeugenaussagen die Wahrscheinlichkeit für p wesentlich mehr als jede einzelne von ihnen⁴⁾. Bedingung dafür ist allerdings, daß die beiden Zeugenaussagen kausal voneinander unabhängig sind. Falls dies erfüllt ist, dann "müssen die Zeugenaussagen sich gegenseitig einen Grad an 'Korroboration' geben".

2. Zwei Indizien eines Indizienbeweises konvergieren, wenn sie, unabhängig voneinander, ein und dieselbe Schlußfolgerung wahrscheinlicher machen. Angenommen, es ist mit an Sicherheit grenzender Wahrscheinlichkeit festgestellt, daß A durch das Testament von B den letzteren beerben soll, so daß A ein Motiv für die Ermordung von B hatte; diese Tatsache erhöht die Wahrscheinlichkeit, daß A B getötet hat. Ähnliches gilt, falls festgestellt ist, daß A zur Tatzeit bei B zu Gast war, so daß er die Möglichkeit zur Tat hatte; das erhöht auch die Wahrscheinlichkeit, daß A B getötet hat. Zusammen erhöhen diese beiden Tatsachen die Wahrscheinlichkeit dafür, daß A B getötet hat, aber wesentlich stärker als jede einzelne für sich allein. (Die Voraussetzung ist hier, daß die beiden Indizien voneinander unabhängig sind).

Die logische Struktur der Korroboration und Konvergenz ist somit folgende⁵⁾
Wenn $P(S/R_1) > P(S)$ und $P(S/R_2) > P(S)$ und R_1 und R_2 voneinander unabhängig sind, dann gilt: $P(S/R_1 \& R_2) > P(S/R_1)$ und $P(S/R_1 \& R_2) > P(S/R_2)$.

1) Cohen: "The Difficulty about Corroboration and Convergence", S.91.
2) *ibid*, S.94f.

3) *ibid*, S.94: "On its own the admitted fact that A testifies that p would do something to raise the probability that p, and so would the fact that B testifies that p". Diese Behauptung ist insofern nicht korrekt, als hier der Tatsache nicht Rechnung getragen wird, daß auch dann, wenn ein Zeuge zugunsten einer Proposition aussagt, das die Wahrscheinlichkeit für die Wahrheit dieser Proposition nicht erhöhen muß, etwa dann, wenn der Zeuge uns nicht glaubwürdig erscheint. So kann ein Zeuge (in Cohens Beispiel auf S.96) angeben, daß der Angeklagte in seiner Gegenwart die Tat angekündigt hat; diese Aussage spricht also zugunsten der Annahme, daß der Angeklagte der Täter ist. Die Wahrscheinlichkeit der Annahme aufgrund der Zeugenaussage muß aber deswegen nicht höher sein als vorher; sie kann sogar kleiner werden, etwa, wenn irgendwelche Gründe vorliegen, zu glauben, daß die Wahrscheinlichkeit dafür, daß der Zeuge auf seiner Aussage besteht, wenn die Annahme falsch ist, größer ist, als wenn diese richtig ist.

4) Auch hier gilt die obige Fußnote; außerdem kann man genug Fälle finden, wo jede einzelne Zeugenaussage die Wahrscheinlichkeit der Proposition erhöht, alle zusammen aber ihre Wahrheit unmöglich machen.

5) Vgl. Cohen, S.94f.

Im folgenden wird untersucht, ob wir in der Tat, wie Cohen behauptet, Schwierigkeiten haben, solche Fälle wahrrscheinlichkeitstheoretisch zu behandeln.

Zunächst überprüfen wir, ob die Relation in einem korroborierenden Zeugenbeweis bzw. in einem konvergierenden Indizienbeweis, daß $P(S/R_1) > P(S)$ und $P(S/R_2) > P(S)$ ist, immer gegeben ist, wenn R_1 bzw. R_2 zugunsten von S sprechen (bei Cohen ist das nicht klar). Dieser Nachweis scheint angesichts der Fußnote 3 auf der vorigen Seite überflüssig zu sein. Wir wollen trotzdem eine formale Rekonstruktion dieser informellen Argumentation an einem Beispiel geben. Wir beschränken uns dabei auf den Zeugenbeweis, denn die Struktur der beiden Beweise, um die es hier geht, ist dieselbe.

Wir nehmen an, daß zwei Zeugen A und B ihre Aussagen vor Gericht in einem Fall machen, in dem es darum geht, ob X Y getötet hat oder nicht. A sagt aus, daß er unmittelbar erlebt hat, wie X Y getötet hat, und B sagt aus, daß er X zur Tatzeit vom Tatort weglaufen gesehen hat. Die Aussage von A und B (wir kürzen sie mit R_1 und R_2 ab) sprechen also zugunsten der (oder bestärken die) Hypothese, daß X Y getötet hat (wir kürzen diese Hypothese mit S ab). Wir wollen also wissen, ob $P(S/R_1) > P(S)$ und $P(S/R_2) > P(S)$ ist. Da es sich um bedingte Wahrscheinlichkeiten handelt, ist die Formel für die bedingte Wahrscheinlichkeit oder das Bayes-Theorem anzuwenden. Wir wenden das letztere an.

Zur Vereinfachung nehmen wir (ohne weitere Begründung) folgende Wahrscheinlichkeitswerte an:

$P(S) = 0.1$ (die Anfangswahrscheinlichkeit, daß X Y getötet hat im Lichte anderer Beweismittel als die beiden Zeugenaussagen, z.B., daß X ein Motiv hatte, Y zu töten);

$P(R_1/S) = 1$ (d.h. A wird X auf jeden Fall belasten, wenn X die Tat begangen hat);

$P(R_1/\neg S) = 1$ (d.h. A wird X auch dann auf jeden Fall belasten, wenn X die Tat nicht begangen hat, z.B., weil A X haßt oder von der Verurteilung von X profitiert etc.);

$P(R_2/S) = 0.7$ (d.h. die Wahrscheinlichkeit, daß B X belasten wird, wenn X die Tat begangen hat, ist relativ hoch, jedoch nicht sehr hoch, weil B vielleicht leicht kurzfristig ist oder irgendwelche Skrupel hat, X zu belasten);

$P(R_2/\neg S) = 0.1$ (d.h. die Wahrscheinlichkeit, daß B X belasten wird, wenn X die Tat nicht begangen hat, ist sehr gering; sie ist jedoch nicht ganz Null, weil B leicht kurzfristig ist oder aus sonstigen Gründen).

Mit dem Bayes-Theorem erhalten wir:

$P(S/R_1) = 0.1$ und $P(S/R_2) \approx 0.4$.

Während die erste Zeugenaussage die Anfangswahrscheinlichkeit für S überhaupt nicht verändert, erhöht die zweite Zeugenaussage diese Wahrscheinlichkeit wesentlich. Wir stellen somit fest, daß es überhaupt nicht notwendig ist, daß $P(S/R) > P(S)$ ist, wenn R zugunsten von S spricht (Es sei ohne Demonstration gesagt, daß $P(S/R)$ auch kleiner als $P(S)$ sein kann).

Für unsere nachfolgende Untersuchung nehmen wir jedoch an, daß wir es hier mit einem Fall zu tun haben, wo sowohl $P(S/R_1) > P(S)$ wie auch $P(S/R_2) > P(S)$ erfüllt ist, falls R_1 resp. R_2 zugunsten von S sprechen.

Als nächstes wollen wir zeigen, daß im folgenden Beispiel $P(S/R_1 \& R_2) > P(S/R_1)$ und $P(S/R_1 \& R_2) > P(S/R_2)$ nicht gelten¹⁾.

Angenommen, X wird wegen eines Raubüberfalls angeklagt (Abk. S). Weiter sei angenommen, daß der Raubüberfall zwischen 17.00 und 17.30 stattgefunden und mindestens 20 Minuten gedauert hat. Wir haben wiederum zwei Zeugen: Der Zeuge A sagt aus, daß er den Angeklagten in der Nähe des Tatorts²⁾ um 17.20 gesehen hat (Abk. R_1); der Zeuge B sagt aus, daß er ihn in der Nähe des Tatorts um 17.10 gesehen hat (Abk. R_2).

R_1 wie auch R_2 sprechen jede für sich zugunsten der Hypothese, daß der Angeklagte den Raubüberfall verübt hat. Wir nehmen weiter an, daß die beiden Zeugen sehr glaubwürdig sind, so daß $P(S/R_1) > P(S)$ und $P(S/R_2) > P(S)$ gilt.

Trotzdem erhalten wir aber $P(S/R_1 \& R_2) = 0$, denn wenn beide Zeugenaussagen stimmen, d.h. der Angeklagte sowohl um 17.10 wie auch um 17.20 in der Nähe des Tatorts gesehen wurde, dann kann er ja den Raubüberfall, der nach der obigen Annahme mindestens 20 Minuten gedauert hat, nicht begangen haben³⁾.

Nun kehren wir zu unserem Beispiel auf S.205 zurück. Wir wollen zeigen, daß im besprochenen Fall (mit einer günstigen Wahl von Wahrscheinlichkeitswerten) $P(S/R_1 \& R_2) > P(S/R_1)$ und $P(S/R_1 \& R_2) > P(S/R_2)$ ist (wobei wir annehmen, daß R_1 und R_2 jede für sich zugunsten der Annahme sprechen und außerdem ihre Wahrscheinlichkeit erhöhen, d.h. es gilt $P(S/R_1) > P(S)$ und $P(S/R_2) > P(S)$). Später geben wir einen allgemeinen Beweis für diese Behauptung.

Zur Vereinfachung nehmen wir folgende Wahrscheinlichkeitswerte an⁴⁾

$P(S) = 0.1$ (wie vorher);

$P(S/R_1) = 0.5$ (etwa, wenn $P(R_1/S) = 1$ und $P(R_1/\neg S) = 0.1$;

$P(S/R_2) = 0.4$ (wie oben).

Ferner nehmen wir an, daß die Zeugen (-aussagen) voneinander bedingt unabhängig sind. Es gilt somit: $P(R_1 \& R_2/S) = P(R_1/S)P(R_2/S)$ und entsprechend $P(R_1 \& R_2/\neg S) = P(R_1/\neg S)P(R_2/\neg S)$.

Mit dem Bayes-Theorem

$$\begin{aligned} P(S/R_1 \& R_2) &= \frac{P(S)(R_1 \& R_2/S)}{P(S)P(R_1 \& R_2/S) + P(\neg S)P(R_1 \& R_2/\neg S)} \\ &= \frac{P(S)P(R_1/S)P(S/R_2)}{P(S)P(R_1/S)P(R_2/S) + P(\neg S)P(R_1/\neg S)P(R_2/\neg S)} \end{aligned}$$

1) dies ist auch bereits auf S.204 Fn.4. erwähnt.

2) aber nicht direkt am Tatort.

3) Dies ist sowohl ein intuitiv plausibles wie auch wahrscheinlichkeitstheoretisch nachvollziehbares Ergebnis. Erstaunlich ist, daß Tribe (1971), der auf S.1367 ein ähnliches Beispiel angibt, dies anscheinend als Gegenargument gegen die Wahrscheinlichkeitstheorie benutzt.

4) die Zahlen werden abgerundet.

erhalten wir:

$$P(S/R_1 \& R_2) = \frac{0.1 \cdot 1 \cdot 0.7}{0.1 \cdot 1 \cdot 0.7 + 0.9 \cdot 0.1 \cdot 0.1} = 0.89.$$

Es gilt somit (in diesem Beispiel):

$P(S/R_1 \& R_2) > P(S/R_1)$ und $P(S/R_1 \& R_2) > P(S/R_2)$, was zu zeigen war.

Im obigen Beispiel haben wir die gegenseitige Unabhängigkeit von R_1 und R_2 als eine bedingte Unabhängigkeit relativ zu S verstanden. Denn aus der nicht bedingten Unabhängigkeit von R_1 und R_2 folgt nicht, daß sie auch relativ zu irgendwelchen Informationen voneinander unabhängig sind. Für die Konvergenz oder Korroboration im wahrscheinlichkeitstheoretischen Sinn ist aber die letzte Art der der Unabhängigkeit wichtig.

Wir wollen nun allgemein beweisen, daß unter bestimmten Voraussetzungen stets $P(S/R_1 \& R_2) > P(S/R_1)$ und $P(S/R_1 \& R_2) > P(S/R_2)$ gilt.

Diese Voraussetzungen sind:

- (i) $P(R_1/S) > P(R_1/\neg S)$ und $P(R_2/S) > P(R_2/\neg S)$;
- (ii) $P(R_1 \& R_2/S) = P(R_1/S)P(R_2/S)$ und $P(R_1 \& R_2/\neg S) = P(R_1/\neg S)P(R_2/\neg S)$ (d.h. R_1 und R_2 sind voneinander bedingt (durch S und $\neg S$) unabhängig)¹⁾

Wir benutzen folgende Abkürzungen:

$$P(S/R_1 \& R_2) = x; P(S/R_1) = a; P(S/R_2) = b; P(S) = s; P(R_1/S) = y; P(R_2/S) = w; \\ P(R_1/\neg S) = v; P(R_2/\neg S) = z.$$

Nach Voraussetzung gilt:

- (i) $y > v$ und $w > z$;
- (ii) $P(R_1 \& R_2/S) = P(R_1/S)P(R_2/S) = y \cdot w$ und $P(R_1 \& R_2/\neg S) = P(R_1/\neg S)P(R_2/\neg S) = v \cdot z$.

Zu zeigen ist: i) $x > a$ und ii) $x > b$.

Beweis:

$$i) \quad \frac{s \cdot y \cdot w}{s \cdot y \cdot w + (1-s) \cdot v \cdot w} = \frac{s \cdot y}{s \cdot y + (1-s) \cdot v} = a;$$

zu zeigen ist:

$$x = \frac{s \cdot y \cdot w}{s \cdot y \cdot w + (1-s) \cdot v \cdot z} > \frac{s \cdot y \cdot w}{s \cdot y \cdot w + (1-s) \cdot v \cdot w}, \text{ d.h.}$$

$$\frac{w}{w + z} > \frac{w}{w + v}$$

Da nach Voraussetzung $w > z$ und außerdem $w, z \in [0,1]$ gilt, ist dies trivialerweise erfüllt. Somit erhalten wir $x > a$.

ii) Beweis $x > b$ analog.

(Analog läßt sich zeigen, daß aus $y > v$ und $w > z$ folgt: $a > s$ und $b > s$).

¹⁾ Vgl. dazu die (mir unverständliche) Argumentation Cohens gegen eine ansatzweise ähnliche Rekonstruktion der Korroboration-Konvergenz bei Reichenbach, *The Direction of Time* (1956) S.157ff; Cohen, S.103f, Fn.6.

Wir sehen also, daß die Kritik Cohens an der Wahrscheinlichkeitstheorie, man könne bestimmte juristische Argumentationen nicht wahrscheinlichkeits-theoretisch nachvollziehen, nicht zutrifft. Unsere Analyse hat gezeigt, daß der von Cohen angegebene Fall der Zeugenkorroborations (und entsprechend der der gleich strukturierten Indizienkonvergenz) mit der Bayes-Methode lösbar ist¹⁾. Außerdem hat diese Analyse gezeigt, daß nicht immer Zeugenkorroborations bzw. Indizienkonvergenz gilt, auch dann, wenn jede einzelne Zeugenaussage (jedes einzelne Indiz) die Sachverhaltsannahme wahrscheinlicher macht (was sowohl intuitiv plausibel wie auch wahrscheinlichkeits-theoretisch nachvollziehbar ist).

Wir wollen hier noch auf die Kritik Cohens hinsichtlich der Apriori-Wahrscheinlichkeiten kurz eingehen. Diese sind nach Cohen in der Rechtspraxis unzulässig. Der Grund ist, wir zitieren (S.108):

"...in the Anglo-American system of procedure an accused person is to be judged only on the facts before the court. Hence he does not come into court with a certain positive prior probability of guilt".

Cohen vertritt irrtümlicherweise die Auffassung, daß die Apriori-Wahrscheinlichkeitsbeurteilung einer Sachverhaltshypothese nichts mit den Tatsachen zu tun hat, die für den juristischen Prozeß relevant sind. Wäre das der Fall, so widerspräche das in der Tat den juristischen Anforderungen. Allerdings haben bereits Finkelstein & Fairley gezeigt, wie man die Apriori-Wahrscheinlichkeiten interpretieren muß, um eine mit der Rechtspraxis konsistente Beurteilung zu erhalten: Sie haben der Hypothese einen (Apriori-) Wahrscheinlichkeitswert im Lichte aller nichtstatistischen Daten zugeordnet. Auch in dieser Arbeit wurde bereits darauf eingegangen, was man unter Apriori- (bzw. Anfangs-) Wahrscheinlichkeiten verstehen soll²⁾.

1) Cohen erwähnt mit keinem Wort die Bayes-Methode, obwohl sie nicht nur in der Wissenschaftstheorie und Statistik, sondern spätestens seit Finkelstein & Fairley (1970) auch in der juristischen Fachliteratur diskutiert wird. Anstatt dessen befaßt er sich ausführlich mit zwei für die Lösung solcher Fälle nicht adäquaten Methoden: eine stammt von Boole und die andere von Ekelöf (vgl. Cohen, S.95ff). Vgl. meine Kritik an diesen beiden Methoden S.87, 88ff, 109f (Auch Cohen lehnt diese Methoden als ungeeignet ab). Allerdings kommt auch Cohen (auf S.101-107) - durch eine ziemlich umständliche algebraische Analyse - zur Auffassung, daß die Wahrscheinlichkeitstheorie doch zur adäquaten Analyse der Zeugenkorroborations und der Indizienkonvergenz führt.

2) Vgl. S.136f. dieser Arbeit.

Diese sind nichts anderes als subjektive Schätzungen der Sachverhaltshypothese im Lichte bestimmter, für den Prozeß relevanter Informationen. Es wurde dabei gezeigt, daß ein Apriori-Wahrscheinlichkeitsurteil sogar alle vor Gericht gebrachten relevanten Informationen erhalten kann¹⁾; in diesem Fall ist der Unterschied zwischen dem qualitativen (üblichen) und dem Wahrscheinlichkeitsurteil nur, daß ersterem numerische Werte zugeordnet werden. Ein solches Urteil (das eigentlich dann auch ein Aposteriori-Urteil ist) ist natürlich noch zu vage. Will man eine präzisere und konsistentere Beurteilung der Sachverhaltshypothese, so wird man das subjektive Apriori-Urteil nicht im Lichte aller, sondern relativ weniger relevanter Informationen machen (wobei es praktisch irrelevant ist, welche das sein sollen), um aufgrund der restlichen Informationen die Aposteriori-Wahrscheinlichkeit abzuleiten.

Cohen führt übrigens auch eine weitere Kritik des Apriori-Ansatzes an. Auf S.107 schreibt er, daß auch dann, wenn eine oder mehrere Komponenten einer zusammengesetzten Hypothese eine Aposteriori-Wahrscheinlichkeit erhalten, die kleiner ist als die Apriori-Wahrscheinlichkeit, die Aposteriori-Wahrscheinlichkeit für die zusammengesetzte Hypothese größer sein kann als die Apriori-Wahrscheinlichkeit für diese. Das widerspricht jedoch der Rechtspraxis, wo jede Komponente einer zusammengesetzten Hypothese aufgrund von Daten mit überwiegender Wahrscheinlichkeit (im Sinn > 0.5) festgestellt werden muß (vgl. auch Cohen S.64, wo er ausführlicher ist).

Wir illustrieren Cohens Kritik an einem Beispiel:

Angenommen, wir haben zwei einfache, voneinander unabhängige Hypothesen, H_1 und H_2 mit $P(H_1) = 0.7$ und $P(H_2) = 0.3$ und ein relevantes Datum D , so daß $P(H_1/D) = 0.6$ und $P(H_2/D) = 0.4$ gilt. Es gilt somit: $P(H_1/D) < P(H_1)$, $P(H_2/D) > P(H_2)$, aber $0.24 = P(H_1 \& H_2/D) > P(H_1 \& H_2) = 0.21$.

Die Lösung ist hier analog zu der ersten Cohenschen Paradoxie (siehe S.172ff dieser Arbeit). Man vergleicht die Wahrscheinlichkeiten $P(H_1 \& H_2/D) = 0.24$, $P(\neg H_1 \& H_2/D) = 0.16$, $P(H_1 \& \neg H_2/D) = 0.36$ und $P(\neg H_1 \& \neg H_2/D) = 0.24$. Wie wir sehen, überwiegt $P(H_1 \& \neg H_2/D)$ (d.h. der Fall, wo die zweite Komponente falsch ist) jede andere Möglichkeit und nicht $P(H_1 \& H_2/D)$.

Zusammenfassung

Die Behauptung Cohens, daß es sehr schwierig (bzw. unmöglich) sei, einen Beweis im einen juristischen Fall wahrscheinlichkeitstheoretisch nachzuvollziehen, in dem entweder korroborierende (sich gegenseitig bestärkende) Zeugenaussagen oder konvergierende Indizien vorliegen, hat sich als unbegründet erwiesen. Wie gezeigt wurde, kann man solche Argumentationen ohne Schwierigkeiten mit Hilfe der Bayes-Methode rekonstruieren.

Der Apriori-Ansatz verlangt, daß u.a. Apriori-Schätzungen der Sach-

1) Darin bestand, wie wir gesehen haben, die Methode von Simon & Mahan, S.150ff dieser Arbeit.

verhaltenshypothese vorgenommen werden. Um diesen Ansatz daher in Einklang mit der juristischen Forderung zu bringen, die die Beurteilung aufgrund anderer als vor Gericht gebrachten Tatsachen verbietet, müssen solche Apriori-Schätzungen im Lichte der für den Prozeß relevanten Informationen erfolgen. Welche und wieviele relevante Informationen zur Bildung des Apriori-Wahrscheinlichkeitsurteils herangezogen werden, entscheidet das Individuum.

Zusammenfassung aller Ergebnisse aus den Abschnitten 1 bis 6:

(die ausführlichen Zusammenfassungen zu den einzelnen Paradoxien und ihren Lösungen findet der Leser am Ende der jeweiligen Abschnitte).

Cohens Anspruch, die Nichtanwendbarkeit der Wahrscheinlichkeitstheorie auf die induktiven Argumentationen in der Rechtspraxis zu zeigen, indem er 6 Paradoxien gegen diese aufstellte, können wir als unerfüllt ansehen, denn diese 'Paradoxien' lassen sich relativ leicht lösen. Wir führen noch einmal auf, worauf Cohens Mißverständnisse in bezug auf die Wahrscheinlichkeitstheorie beruhen:

1. Paradoxie: Eine unkorrekte Rekonstruktion des Möglichkeitsraumes der in Frage kommenden Sachverhaltshypothese und die unzulässige Anwendung des Multiplikationsprinzips in Fällen, wo das Bayes-Theorems (bzw. die Jeffrey-, bzw. die Bernoulli-Formel) anzuwenden sind, bzw. eine fehlende Unterscheidung zwischen einer zusammengesetzten Sachverhaltshypothese und einem zusammengesetzten Datum (mehreren Beweismitteln).

2. Paradoxie: Eine nichtkorrekte Forderung, daß die Wahrscheinlichkeitstheorie irgendwelche Beweiskriterien formulieren und begründen soll (Die Wahrscheinlichkeitstheorie liefert nur einen Rahmen, innerhalb dessen man von bestimmten Annahmen zu anderen Annahmen in konsistenter und konsequenter Weise übergehen kann; die Aufstellung und Begründung von Beweis- bzw. Entscheidungskriterien, z.B. juristischer Art, gehört in einen anderen Kontext - etwa den einer juristischen Entscheidungstheorie) und die falsche Behauptung, daß mehrstufige Beweise vom Gericht nicht zugelassen werden (in diesem Fall müßte man z.B. den Zeugenbeweis zur Feststellung von Beweismitteln (die ihrerseits zum Nachweis der Sachverhaltshypothese verwendet werden) - als typischen Vertreter dieser Kategorie ausschließen).

3. Paradoxie: Eine nicht korrekte Anknüpfung an die statistische Interpretation des Wahrscheinlichkeitsbegriffs bei der Beurteilung von Einzelfällen und eine fehlende (bzw. nicht adäquate) Berücksichtigung der subjektivistischen Interpretation (die nichts mit der

1) Vgl. dazu auch die Kritik von Kaye (S.41): "It seems more sensible (though less original) to say that the probabilities behave as prescribed by the conventional probability axioms, but that the law accepts imprecision and inaccuracy because other values are at work. The putative anomalies Cohen depicts thus do not compel us to abandon the established theory of probability at the courthouse door".

relativen Häufigkeit zu tun hat¹⁾), für die das Problem der Behandlung von Einzelfällen - diese können alltäglich oder juristisch sein - nicht auftaucht.

4.Paradoxie: Eine nicht adäquate Interpretation der mathematischen Wahrscheinlichkeitstheorie bei der Behandlung des Beweiskriteriums "mit an Sicherheit grenzender Wahrscheinlichkeit": Cohen Kritik richtet sich gegen die objektive (statistische) Wahrscheinlichkeitstheorie; bei dem Begriff "mit an Sicherheit grenzender Wahrscheinlichkeit" handelt es sich aber um ein subjektives Beweiskriterium, das daher nur im Rahmen der subjektiven Wahrscheinlichkeitstheorie zu behandeln ist.

5.Paradoxie: Die Nichtberücksichtigung dessen, daß es sich bei der Rechtspraxis um einen Bereich handelt, wo man entscheiden bzw. handeln muß; eine Theorie muß daher die Rationalität solcher Entscheidungen garantieren. Eine solche Theorie ist die Entscheidungstheorie. Der subjektive Wahrscheinlichkeitsbegriff - und nur er - erlaubt somit eine entscheidungstheoretische Rechtfertigung, weil er Grundbegriff dieser Theorie ist.

6.Paradoxie: Die Nichtberücksichtigung der Bayes-Methode, mit der man relativ einfach den korroborierenden Zeugenbeweis und den konvergierenden Indizienbeweis wahrscheinlichkeitstheoretisch rekonstruieren kann. Außerdem die unbegründete Annahme, daß die Apriori-Wahrscheinlichkeitsbeurteilung einer Sachverhaltshypothese im Lichte irgendwelcher für den juristischen Fall relevanten Tatsachen nicht durchführbar sei²⁾³⁾

1) Cohen nimmt fälschlicherweise an, daß der subjektive Wahrscheinlichkeitsbegriff etwas mit relativen Häufigkeiten zu tun hat; vgl.S.192Fn.3 dieser Arbeit.

2) Das ursprüngliche Vorhaben, auch eine kritische Auseinandersetzung mit der Theorie Cohens zu geben, kann, da sie sowohl den Rahmen wie auch den Umfang dieser Arbeit sprengen würde, hier nicht mehr durchgeführt werden.

3) Zur weiteren Kritik an Cohen vgl.: G.Williams, *The Mathematics of Proof*, in: *Crim.Law Rev.* (1979), S.297-308, 340-354; G.Williams, *A Short Rejoinder*, in: *Crim.Law Rev.* (1980), S.103-107; R.Eggleston, *The Probability Debate*, in: *Crim.Law Rev.* (1980), S.678-688; R.O. Havery, *Baconian and Pascalian Probability in Forensic Proof*, *ibid.*, S.743-747; S.Blackburn, *Review of Cohen, The Probable and the Provable*, in: 44 *Synthese* (1980), S.149-159; I.Levi, *Support and Surprise: L.J.Cohen's View of Inductive Probability*, in: 30 *Brit.J. Sci.* (1979), S.279-292; D.Kahnemann & A.Tversky, *On the interpretation of intuitive probability: A reply to Jonathan Cohen*, in: 7 *Cognition* (1979), S.409ff; R.Birmingham, *Remarks on 'Probability' in Law: Mostly, A Casenote and a Book Review*, in: 12 *Georgia Law Rev.* (1978), S.535-551; D.Kaye, *The Laws of Probability and The Law of the Land*, in: 47 *The Univ. of Chicago Law Rev.* (1979), S.34-56; D.A.Schum, *A review of a Case Against Blaise Pascal and His Heirs*, in: 77 *Michigan Law Rev.* (1979), S.446-483.

§4. Wahrscheinlichkeitstheoretische Rekonstruktion juristischer Fälle

In diesem Paragraphen werden zwei juristische Fälle - People v. Collins und ein fiktiver Fall - mit Hilfe von vier für die Rechtspraxis wichtigen Formeln: der Bayes-, der Bernoulli-, der Jeffrey-Formel und dem Multiplikationsprinzip - wahrscheinlichkeits-theoretisch rekonstruiert. Das Ziel ist somit zu zeigen, wie man die subjektive Wahrscheinlichkeitstheorie in der Rechtspraxis korrekt anwendet. Die wahrscheinlichkeits-theoretische Rekonstruktion erfolgt Schritt für Schritt, wobei einige Forderungen an den Anwendungsbereich formuliert, der Begriff der Relevanz in Anlehnung an Gärdenfors definiert und Begriffe wie "Apriori-Wahrscheinlichkeit", "Zeugenglaubwürdigkeit" usw. erläutert werden.

Zunächst wollen wir einige (plausibel erscheinende) Forderungen an den Anwendungsbereich stellen. Diese sind:

(1) Die Forderung des zulässigen Datums:

Ein juristischer Sachverhalt (oder genauer: eine juristische Sachverhaltshypothese) muß aufgrund der vor Gericht zugelassenen Daten beurteilt werden.

(2) Die Forderung des relevanten Datums:

Ein juristischer Sachverhalt muß aufgrund der für ihn relevanten Daten beurteilt werden*}

Welches Datum relevant bzw. irrelevant ist, entscheidet das Individuum**}

Die Definition der Relevanz:

Die gewöhnliche Definition der Relevanz lautet wie folgt¹⁾

Ein Datum D_i ist genau dann für die Hypothese H relevant, wenn $P(H/E \wedge D_i) \neq P(H/E)$ ist. E ist dabei das allgemeine Hintergrunddatum. Üblicherweise wird auf die explizite Anführung von E verzichtet, so daß die Definition der Relevanz einfach lautet:

D_i ist relevant für H genau dann, wenn $P(H/D_i) \neq P(H)$.

Entsprechend ist die Definition der Irrelevanz:

D_i ist irrelevant für H genau dann, wenn $P(H/E \wedge D_i) = P(H)$ oder $E \wedge D_i$ ist logisch falsch.

Es kann allerdings der Fall eintreten, daß zwar $P(H/D_i) = P(H)$ ist, dagegen $P(H/D_i \wedge D_j) \neq P(H/D_j)$, $j \neq i$, $j = 1, \dots, n$, d.h. ein bestimmtes Datum kann für sich betrachtet irrelevant für die Hypothese sein, in Verbindung mit einem anderen Datum dagegen relevant. Der umgekehrte Fall ist, daß zwar $P(H/D_i) \neq P(H)$ gilt, aber $P(H/D_i \wedge D_j)$

*) Es sei denn, es verstößt gegen 1.

**) Es sei denn, es verstößt gegen die logischen oder Naturgesetze. Außerdem muß es der folgenden Definition genügen.

1) Vgl. R. Carnap, Logical Foundations of Probability (1950), S. 348; P. Gärdenfors, On the Logic of Relevance, in: 37 Synthese (1978), S. 351-367; S. 351.

= $P(H/D_j)$ bzw. $P(H/D_i \wedge D_j) = P(H)$, d.h. ein Datum, das - für sich betrachtet - für die Hypothese relevant ist, kann trotzdem in Verbindung mit anderen Daten für die Hypothese irrelevant sein. Solche Fälle werden von der obigen Definition nicht erfüllt.

Der zweite Fall ist allerdings weniger unangenehm als der erste: Denn während wir im ersten das ausschalten, was sich im nachhinein als relevant, d.h. wahrscheinlichkeitsverändernd, erweist, berücksichtigen wir im letzteren etwas, was sich später in Verbindung mit einem anderen Datum als irrelevant herausstellt und daher (in Verbindung mit diesem Datum) die ursprüngliche Wahrscheinlichkeitsbewertung der Hypothese nicht verändert bzw. wiederherstellt¹⁾.

Wir wollen dies bei der nachfolgenden Definition der Relevanz berücksichtigen. Außerdem wollen wir berücksichtigen, daß es sich bei der Gesamtheit der auf ihre Relevanz beurteilten juristischen Daten um eine endliche Anzahl bereits festgestellter bzw. zur Verfügung stehender Daten handelt. Ferner, daß die Relevanz eines bestimmten Datums im Lichte des gesamten festgestellten Datums beurteilt werden soll.

Eine allgemein akzeptierte Definition der Relevanz, die den obigen Fällen gerecht wird, existiert meines Erachtens nicht; vgl. dazu auch Gärdenfors (S.351). Auf S.362 gibt Gärdenfors folgende Definition der Relevanz an:

(a) p is irrelevant to r on evidence e iff either (i) $P(r/p \& e) = P(r/e)$ and for all sentences q , if $P(r/q \& e) = P(r/e)$ and $P(p \& q \& e) \neq 0$, then $P(r/p \& q \& e) = P(r/e)$, or (ii) $P(p/e) = 0$; (b) p is relevant to r on e iff p is not irrelevant to r on e .

Diese Definition scheint aus folgenden Gründen allerdings modifizierungsbedürftig zu sein:

- 1) Es kann trotzdem ein q' geben, so daß zwar die Bedingungen unter (i) in bezug auf p, q und q' erfüllt sind (d.h. $P(r/p \& e) = P(r/e)$ und für alle Sätze q, q' , wenn $P(r/q \& e) = P(r/e)$ und $P(r/q' \& e) = P(r/e)$ und $P(p \& q \& e) \neq 0$ und $P(p \& q' \& e) \neq 0$, dann $P(r/p \& q \& e) = P(r/e)$ und $P(r/p \& q' \& e) = P(r/e)$), aber $P(r/p \& q \& q' \& e) \neq P(r/e)$ ist. Statt "für alle Sätze q " muß daher heißen "für alle Sätze q und alle möglichen Konjunktionen von Sätzen q ";
- 2) Es werden hier solche Fälle nicht berücksichtigt, wo man p auch dann intuitiv für irrelevant hält, wenn man es mit einem für die Hypothese relevanten Datum verbindet, d.h. daß zwar $P(r/p \& e) = P(r/e)$ und $P(r/q \& e) \neq P(r/e)$ ist, aber $P(r/p \& q \& e) = P(r/q \& e)$.

(Gärdenfors Definition leidet z.B. darunter, daß auch dann, wenn wir kein q mit $P(r/q \& e) = P(r/e)$ finden (d.h. alle Sätze q sind relevant für r), wir trotzdem p als 'irrelevant' ausschalten müssen, obwohl $P(r/p \& q \& e) \neq P(r/q \& e)$ sein kann).

Wir benutzen folgende Abkürzungen²⁾

D_i : für das Datum, das auf seine Relevanz beurteilt wird;

D : für das gesamte bei der Beurteilung der juristischen Hypothese H zur Verfügung stehende (bzw. festgestellte) Datum $D_1, \dots$

D_n ;

1) Auch der zweite Fall kann unangenehm werden, wenn D_j seinerseits vorher als für H irrelevant ausscheidet (siehe 1. Fall).

2) Es wird auf die explizite Anführung von E verzichtet.

D^* : für eine beliebige Kombination (Konjunktion) von Daten aus D ausgenommen D_1 ¹⁾²⁾

Unsere Definition lautet:

Ein bestimmtes Datum D_i ist für die Hypothese H im Lichte des Datums D juristisch irrelevant genau dann, wenn

- (1) $P(H/D_i) = P(H)$ und
- (2) $AD^*(\text{wenn } P(H/D^*) = P(H), \text{ dann } P(H/D^* \wedge D_i) = P(H))$ und
- (3) $AD^*(P(H/D^* \wedge D_i) = P(H/D^*))$ ³⁾

D_i ist relevant für H genau dann, wenn D_i nicht irrelevant für H ist.

Weitere Forderungen:

(3) Die Forderung des Gesamtdatums:

Ein juristischer Sachverhalt muß aufgrund aller (festgestellten bzw. zur Verfügung stehenden) relevanten Daten beurteilt werden^{*}

(4) Die Forderung des hinreichenden Datums:

Bei der Beurteilung eines juristischen Sachverhalts muß ein möglichst umfangreiches, für die Lösung hinreichendes Datum vorliegen.

Wann das Datum für die Lösung des juristischen Falls hinreichend ist, entscheidet das Individuum.

(5) Die Forderung des eindeutig interpretierbaren Datums:

Ein juristischer Sachverhalt darf nicht aufgrund eines Datums beurteilt werden, das je nachdem, wie man es interpretiert, sowohl zugunsten des Sachverhalts wie auch zugunsten seiner Negation spricht.

1) Zur Veranschaulichung: Seien drei Daten D_1, D_2 und D_3 gegeben. Ferner nehmen wir an, daß das Datum D_1 auf seine Relevanz überprüft wird. Dann kann D^* so aussehen: D_2 , aber auch so: D_3 , aber auch so: $D_2 \wedge D_3$.

2) Im folgenden setzen wir voraus, daß für alle D_j , $j = 1, \dots, n$, $P(D_j) \neq 0$ und für alle D^* , $P(D^*) \neq 0$ und $P(D^* \wedge D_i) \neq 0$ ist.

3) Die Bedingungen (2) und (3) lassen sich zu einer einzigen Bedingung (3) zusammenziehen. Die Bedingung (2) ist die explizite Darstellung des Falls, in welchem ein weiteres für die Hypothese irrelevantes Datum vorhanden ist. (aus (3) folgt außerdem (1) für ein tautologisches D^*).

*) es sei denn, es verstößt gegen 1.

Folgende Begriffe sind für die Rekonstruktion wichtig:

1. Die Sachverhaltshypothese

Darunter wird eine juristische Hypothese der Art: "X hat Y getötet", "X hat den Diebstahl begangen", "X ist nicht der Mörder von Y" usw. oder allgemeiner: "X hat die (Straf-) Tat begangen", "X ist der Täter", "X ist nicht der Täter" verstanden.

2. Das Datum

Darunter werden Beweismittel, Informationen, Tatsachen, Aussagen von Zeugen und dergl. verstanden, die für die Beurteilung einer juristischen Sachverhaltshypothese in Frage kommen.

3. Die Apriori-Wahrscheinlichkeitsbeurteilung

Die Apriori-Wahrscheinlichkeitsbeurteilung der juristischen Sachverhaltshypothese muß im Lichte eines für die Hypothese relevanten Datums erfolgen.

Erfolgt die Apriori-Wahrscheinlichkeitsbeurteilung unter Berücksichtigung aller relevanten Daten, dann ist ein solches Urteil (das gleichzeitig ein Aposteriori-Wahrscheinlichkeitsurteil ist) nicht wesentlich präziser als das traditionelle qualitative Urteil (Der Unterschied zwischen beiden ist nur der, daß beim quantitativen Wahrscheinlichkeitsurteil dem qualitativen (üblichen) Endurteil numerische Werte zugeordnet werden). Will man eine präzisere Beurteilung der Hypothese, so muß man das subjektive Apriori-Wahrscheinlichkeitsurteil im Lichte nicht aller, sondern relativ weniger relevanter Daten machen, um aufgrund der restlichen Daten die Aposteriori-Wahrscheinlichkeit der Hypothese abzuleiten.

4. Zeugenglaubwürdigkeit und Tatsachenzuverlässigkeit

Darunter wird die Wahrscheinlichkeit der Daten (Zeugenaussagen, Tatsachen usw.) verstanden, zum einen relativ auf die Wahrheit der Sachverhaltshypothese, zum anderen relativ auf deren Falschheit. (Evtl. wird unter der Datenzuverlässigkeit bzw. -Glaubwürdigkeit auch nur die Wahrscheinlichkeit der Daten verstanden, wie dies bei Anwendung der Jeffrey-Formel der Fall ist).

Wenn die Wahrscheinlichkeit der (zugunsten der Hypothese sprechenden) Daten unter Voraussetzung der Wahrheit der Hypothese hoch, unter Voraussetzung ihrer Negation dagegen gering ist, dann sagen wir, daß die Daten insgesamt zuverlässig bzw. glaubwürdig sind. In allen anderen Fällen bezeichnen wir die (zugunsten der Hypothese sprechenden) Daten in bezug auf die Wahrheit der Hypothese als insgesamt unzuverlässig bzw. unglaubwürdig.

5. Der subjektive Wahrscheinlichkeitsbegriff

Die wahrscheinlichkeitstheoretische Beurteilung einer juristischen Sachverhaltshypothese aufgrund der Daten ist subjektiv.

6. Die statistische Wahrscheinlichkeit

Statistische Wahrscheinlichkeitsurteile sind nur innerhalb der für die Beurteilung der Sachverhaltshypothese herangezogenen Daten zulässig (und zwar nur dann, wenn das Individuum sie für die Beurteilung der Hypothese für relevant hält).

7. Die Methode

Es gibt vier Methoden, die für die wahrscheinlichkeitstheoretische Lösung juristischer Fälle wichtig sind: das Bayes-Theorem, die B_{erw} -Formel (eine Erweiterung des Bayes-Theorems), die Jeffrey-Formel und das Multiplikationsprinzip.

Das Bayes-Theorem wendet man an, wenn eine Sachverhaltshypothese aufgrund von Daten beurteilt werden soll, wobei die Daten ihrerseits sicher sind; sind sie unsicher (d.h. sie sind selber nur mit einer bestimmten Wahrscheinlichkeit festgestellt), so können wir die Unsicherheit bzgl. dieser Daten eliminieren, vorausgesetzt wir finden eine weitere Information, die die ursprünglichen Daten zwar wahrscheinlich macht, selbst aber mit Sicherheit gilt. (In diesem Fall wird die Wahrscheinlichkeit der Hypothese nicht mehr aufgrund der ursprünglichen (unsicheren) Daten, sondern aufgrund des neuen sicheren Datums bestimmt, wobei die alten Daten als der Bestandteil der Beurteilungsgrundlage auch mitberücksichtigt werden (vgl. die Formel auf S.34). Läßt sich die Unsicherheit bzgl. der Daten nicht auf diese Weise eliminieren, so ist die Jeffrey-Formel anzuwenden.

Das Multiplikationsprinzip kommt zur Anwendung, wenn es sich um eine zusammengesetzte Hypothese handelt: dies ist immer dann der Fall, wenn es sich um einen sog. qualifizierten Tatbestand (wie Mord, Diebstahl mit Waffen usw.) handelt.

8. Die Zuordnung von Wahrscheinlichkeitswerten

Wahrscheinlichkeitswerte werden vom Individuum subjektiv zugeordnet. Das Individuum kann dabei intuitiv (aufgrund seiner Erfahrung und Kenntnisse) vorgehen oder, falls es eine größere Präzision anstrebt, eine der (im Anhang angegebenen) Metrisierungsmethoden benutzen.

(Ausgenommen davon sind statistische Wahrscheinlichkeitswerte; diese können, falls das Individuum sie für die Beurteilung der Hypothese für relevant hält, direkt übernommen werden).

Und schließlich wollen wir uns über die Vorgehensweise bei der wahrscheinlichkeitstheoretischen Rekonstruktion eines juristischen Falls Klarheit verschaffen.

Die Vorgehensweise:

1. Schritt: Ausformulierung der Sachverhaltshypothese (unter Angabe, ob es sich dabei um eine einfache oder zusammengesetzte Hypothese handelt; im letzten Fall müssen außerdem auch die einfachen Hypothesen, aus denen die zusammengesetzte Sachverhaltshypothese besteht, ausformuliert werden).

- 2.Schritt: Spezifizierung des gesamten vorliegenden Datums.
- 3.Schritt: Spezifizierung des gesamtrelevanten und Aussonderung des irrelevanten und die Forderungen 1 und 5 verletzenden Datums.
- 4.Schritt: Spezifizierung der Daten, die für die anderen Daten, denjenigen für die Hypothese und denjenigen für die zusammengesetzte Hypothese, relevant sind.
- 5.Schritt: Spezifizierung des unsicheren Datums.
- 6.Schritt: Überprüfung, ob das vorliegende relevante Datum für die Beurteilung hinreichend ist (wenn nicht, müssen zusätzliche Informationen beschafft werden bzw. die Anklage zurückgewiesen werden, wenn zusätzliche Informationen nicht zu beschaffen sind).
- 7.Schritt: Entscheidung bzgl. der anzuwendenden wahrscheinlichkeitstheoretischen Methode.
- 8.Schritt: Entscheidung darüber, welche Daten man für die Bildung des Apriori-Wahrscheinlichkeitsurteils heranzieht (die restlichen relevanten Daten müssen dann für die Bildung des Aposteriori-Wahrscheinlichkeitsurteils benutzt werden).
- 9.Schritt: Zuordnung von Wahrscheinlichkeitswerten: Entscheidung bzgl. der dabei anzuwendenden Metrisierungsmethode und Begründung der angenommenen subjektiven Wahrscheinlichkeitswerte.
- 10.Schritt: Rechenvorgang.

Rekonstruktion

People v. Collins

(eine ausführliche Beschreibung dieses Falls ist auf S.112ff dieser Arbeit zu finden).

1. Die Sachverhaltshypothese lautet:

Das angeklagte Paar (Janet Collins und Malcolm Collins) hat den Raubüberfall begangen.

Wir kürzen diese Hypothese mit H ab.

Es handelt sich dabei um eine zusammengesetzte Sachverhaltshypothese (die den qualifizierten Tatbestand "Raub" beschreibt¹⁾), die aus zwei einfachen Hypothesen besteht:

- 1) Das angeklagte Paar hat die Tat begangen (Abk. H₁)

und

- 2) Die Tat wurde unter Anwendung von Gewalt oder auf eine andere vom §249 StGB als Raub qualifizierte Weise verübt (Abk. H₂).

2. Das gesamte vorliegende Datum:

(Die Schilderung des Überfalls und anderer Tatsachen, die im Prozeß eine Rolle spielen; vgl.S.112ff dieser Arbeit).

1) Beim "Raub" muß nämlich nicht nur nachgewiesen werden, daß etwas gestohlen wurde, sondern daß dies "mit Gewalt" oder "unter Anwendung von Drohungen mit gegenwärtiger Gefahr für Leib oder Leben" geschah (vgl.§249,Raub.StGB). Wir rekonstruieren also den amerikanischen Rechtsfall im Rahmen deutscher Rechtsverordnungen.

3. Das für die Beurteilung relevante Datum¹⁾

1) Die Schilderung des Täterpaars aufgrund der Zeugenaussagen von Mrs. Juanita Brooks und Mr. John Bass:

J.Brooks beschrieb die Täterin als eine junge Frau, die "etwas schwarzes" trug, ca 145 Pfund wog und "zwischen dunkelblond und hellblond" war.

J.Bass beschrieb die Frau, die aus der Allee, aus der er kurz zuvor Schreie vernommen hatte, hinauslief und in ein gelbes Auto einstieg, als vom kaukasischen Typ, etwas über 1.50 m groß, normal gebaut mit dunkelblondem Haar in Pferdeschwanzfrisur; den Mann, der am Steuer des gelben Autos saß, als von schwarzer Hautfarbe mit Bart und Schnurrbart.

2) Andere Zeugen (als J.Brooks und J.Bass) sagten aus, daß Malcolm Collins am Tagtag einen Bart und Schnurrbart trug und einen gelben Lincoln mit weißem Dach fuhr.

3) Andere Zeugen (als J.Brooks und J.Bass) sagten aus, daß Janet Collins am Tagtag eine Pferdeschwanzfrisur trug und außerdem ihren Arbeitsplatz ungefähr zur Zeit des Überfalls verließ.

4) J.Collins versuchte nach dem Tag des Überfalls, ihre Haarfarbe zu verändern.

5) Die Aussage der Polizisten, daß Malcolm Collins, als sie ihn verhaften wollten, durch die Hintertür weglief und sich auf der Toilette des Nachbarn versteckte.

6) Am Tag nach dem Überfall zahlte Malcolm eine Geldstrafe ungefähr vom gleichen Betrag, der gestohlen wurde (zu der bei ihm gefundenen Quittung erklärte er, daß er das Geld in einem Glücksspiel gewonnen habe).

7) Janet Collins machte während des Verhörs Andeutungen, daß sie gestehen würde, wenn man ihren Mann freiläße.

8) J.Bass konnte sowohl Janet wie auch Malcolm Collins einwandfrei vor Gericht identifizieren; außerdem konnte er Janet auch auf einem Polizeiphotographen identifizieren.

9) Mrs. Brooks wurde während des Überfalls von der Täterin zu Boden gestoßen.

10) Die Zeugen des Überfalls²⁾ waren sich uneinig, ob das Auto gelb mit weißem Dach (wie bei dem Angeklagten) oder nur gelb war.

11) Die Zeugen waren sich uneinig, ob Janets Haar dunkler war als das der Täterin (J.Brooks war der Meinung, daß die Täterin helleres Haar hatte).

12) J.Bass hatte bei der ersten Gegenüberstellung kurz nach dem Raubüberfall Schwierigkeiten, Malcolm Collins als den Mann am Steuer des gelben Autos zu identifizieren (was möglicherweise darauf zurückzuführen ist, daß Malcolm Collins zu dieser Zeit keinen Bart trug).

13) Die Angeklagten bestritten die Tat und beteuerten, zur Tatzeit bei einem Freund, der weit vom Tatort entfernt wohne, zu Besuch gewesen zu sein.

14) Der Freund der Angeklagten bestätigte das (er konnte allerdings nicht genau das Datum des Besuchs nennen).

1) Es werden zuerst die zugunsten der Hypothese H_1 sprechenden Daten (1-8), dann das zugunsten der (den Tatbestand "Raub" qualifizierenden) Hypothese H_2 sprechende Datum (9) und schließlich die zugunsten der Hypothese $\neg H_1$ sprechenden Daten (10-14) angeführt.

2) Aus der Schilderung des Falls ist leider nicht ersichtlich, was für Zeugen das waren. Vgl. Cullison, S. 475.

Anzumerken ist folgendes: Solange bestimmte Daten nicht überprüft bzw. nachgewiesen sind, sprechen sie nicht ganz eindeutig gegen die Angeklagten. Aus der Schilderung des Falls geht leider nicht hervor, ob Janet Collins tatsächlich ca 145 Pfund wog, etwas über 1.50 m groß, vom kaukasischen Typ und normal gebaut war; außerdem, ob Malcolm Collins nicht doch andere Gründe hatte, sich vor den Polizisten zu verstecken bzw. ob er nicht doch das Geld in einem Glücksspiel gewonnen hatte. (Hier nehmen wir allerdings an, daß all diese Daten eindeutig gegen Collins sprechen).

Wie wir sehen, sind es insgesamt viel mehr Tatsachen, die für den vorliegenden Fall relevant sind, als bei der auf S.113 dieser Arbeit geschilderten Beurteilung dieses Falls berücksichtigt wurden. Die statistische Beurteilung verstieß somit gegen die Forderung des gesamten relevanten Datums - denn der juristische Sachverhalt wurde nur aufgrund eines Teils des unter 1) angeführten Datums beurteilt (d.i. zum Teil gelbes Auto,¹⁾ Mann mit Schnurrbart, Mädchen mit Pferdeschwanz, Mädchen mit blondem Haar²⁾ schwarzer Mann mit Bart und schwarz-weißes Paar im Auto³⁾). Auch die gerichtliche Entscheidung, das angeklagte Paar freizusprechen, verstieß gegen diese Forderung: denn die relativ hohe Wahrscheinlichkeit von 0.42 (für die Möglichkeit, ein weiteres Paar mit den gleichen Merkmalen aufzufinden⁴⁾), die Anlaß zum Freispruch gab, basierte lediglich auf den oben genannten statistischen Daten.

Ein Beispiel des Verstoßes gegen die Forderung des zulässigen Datums wäre z.B., wenn man eine (im obigen Fall allerdings nicht gegebene) Vorstrafe Malcolms Collins berücksichtigen würde, obwohl die juristische Praxis eine solche Berücksichtigung i.d.R. verbietet.

Gegen die Forderung des relevanten Datums würde man im obigen Fall verstoßen, wenn man etwa die Tatsache berücksichtigt, daß die Geldbörse von J.Brooks auf dem Einkaufswagen lag. Diese Tatsache wäre nur dann relevant, wenn man die Tatsache, daß Mrs. Brooks bestohlen wurde, anzweifeln würde (was hier nicht der Fall ist).

Ein Verstoß gegen die Forderung des hinreichenden Datums wäre z.B., wenn das Gericht zwar das statistische Datum allein für die Beurteilung des Falls für nicht hinreichend hält, seine Entscheidung trotzdem (wie im obigen Fall) darauf stützt (wobei außer den statistischen sonst keine anderen Daten festgestellt sind).

Ein Verstoß gegen die Forderung des eindeutig interpretierbaren Datums wäre z.B. die Berücksichtigung der Tatsache, daß Malcolm Collins kurz nach dem Überfall keinen Bart trug, solange nicht nachgewiesen ist, wann er sich rasiert hat. Denn das kann sowohl zugunsten seiner Täterschaft ausgelegt werden (falls er sich erst nach dem Überfall rasiert hat) oder zugunsten seiner Nichttäterschaft (falls er sich davor rasiert hat). Findet man dagegen heraus, zu welchem Zeitpunkt Malcolm Collins sich rasiert hat, dann ist die obige Tatsache für die Beurteilung der Sachverhaltshypothese relevant

1) Allerdings besagt die Aussage von Bass, daß es sich um ein gelbes Auto handelt; vgl.S.113, Fn.1 dieser Arbeit.

2) ibid.

3) S.113 dieser Arbeit.

4) S.115 dieser Arbeit.

4. Folgende Daten sind für die Hypothese H_1 relevant:

1) - 8), 12) - 14).

Die Daten, die für die anderen Daten relevant sind:

10) für das Datum, daß das Auto (zum Teil) gelb war; 11) für eine bestimmte Haarfarbe von Janet Collins.

Das unter 9) angegebene Datum ist für die Hypothese H_2 (und somit für H) relevant.

5. Folgende Daten sind unsicher:

ob es sich um ein gelbes oder um ein zum Teil gelbes Auto beim Überfall handelte;

ob die Täterin hellere Haarfarbe hatte als Janet Collins.

Außerdem muß man die Tatsache berücksichtigen, daß die Daten nicht zuverlässiger sein können als die Zeugenaussagen, auf denen sie beruhen. Somit ist auch die Zuverlässigkeit der Daten (unabhängig davon, ob die Zeugen über manche Details uneinig sind) im allgemeinen zu berücksichtigen.

6. Wir machen die Annahme, daß das vorliegende Datum für die Beurteilung des Falls hinreichend ist.

7. Da einige Daten unsicher sind (wobei wir ihre Unsicherheit offensichtlich nicht durch die Zurückführung auf sichere Daten eliminieren können) und wir außerdem die Zuverlässigkeit dieser Daten berücksichtigen müssen, ist die Jeffrey-Methode anzuwenden und anschließend das Multiplikationsprinzip, da es sich um eine zusammengesetzte Sachverhaltshypothese handelt.

8. Wir wollen für die Aposteriori-Wahrscheinlichkeitsbeurteilung der Hypothese H_1 (in Analogie zur bekannten Behandlung dieses Falls, S.113 dieser Arbeit) folgende Daten heranziehen (diese bilden einen Teil der unter 1) angeführten Daten):

a) zum Teil gelbes Auto, b) Mann mit Schnurrbart, c) Mädchen mit Pferdeschwanz, d) Mädchen mit (dunkel-)blondem Haar, e) schwarzer Mann mit Bart, f) schwarze-weißes Paar im Auto.

Damit müssen wir das Apriori-Wahrscheinlichkeitsurteil im Lichte folgender Daten machen:

(1) dem Rest der unter 1) angeführten Daten: bei der Täterin handelt es sich um eine junge Frau, die beim Überfall etwas schwarzes trug, ca 145 Pfund wog, etwas über 1.50 m groß, vom kaukasischen Typ und normal gebaut war (Im folgenden wird angenommen, daß diese Beschreibung auf Janet Collins zutrifft);

und

(2) 2) - 8), 12) - 14).

9. Zuordnung von Wahrscheinlichkeitswerten:

Abkürzungen:

H, H_1 und H_2 für die zusammengesetzte Sachverhaltshypothese und die einfachen Hypothesen (wie auf S.217); D_1 für das Datum, das für die Aposteriori-Beurteilung von H_1 und D_2 das Datum, das für eine solche von H_2 herangezogen wird.

D_1 lautet somit: Die Täter wiesen die unter 8. a) - f) genannten Merkmale auf und D_2 : Mrs. Brook wurde von der Täterin zum Boden gestoßen.

Wir nehmen folgende Wahrscheinlichkeitswerte an:

$P(H_1) = 0.9$; Die Begründung kann etwa so lauten: Das Datum gegen das angeklagte Paar (auch ohne Berücksichtigung der unter 8. a) - f) angegebenen Daten) ist überwältigend (fast alle Tatsachen sprechen gegen sie), wir wollen jedoch vorsichtig sein und nehmen einen kleineren Wahrscheinlichkeitswert an, als es der Intuition vielleicht entsprechen würde.

$P*(D_1) = 0.8$; die Begründung kann etwa so lauten: Die Zeugen, deren Aussagen die Daten beschreiben, erscheinen mir sehr glaubwürdig, so daß ich, wenn sie in allen Punkten des Datums übereinstimmten, keinen Zweifel an diesem Datum hätte; da sie aber in zwei Punkten uneinig sind, d.i. ob das Auto gelb oder nur zum Teil gelb und die Haarfarbe der Täterin dunkelblond (wie bei Janet Collins) oder hellblond war, erscheint mir der obige Wahrscheinlichkeitswert von 0.8 für die Zuverlässigkeit des gesamten unter 8 a) - f) angeführten Datums angemessen zu sein.

$P(D_1/H_1) = 1$; die Begründung kann etwa so lauten: Wenn die Angeklagten die Täter waren, dann mußten die Täter (also sie) auch alle im Datum spezifizierten Merkmale aufweisen.

$P(D_1/\neg H) = 5/10^8$; die Begründung kann etwa so lauten: Wenn die Angeklagten nicht die Täter waren, dann ist die Wahrscheinlichkeit daß die Täter die Merkmale a) - f) aufwiesen, gleich der statistischen Wahrscheinlichkeit für das gemeinsame Auftreten dieser Merkmale in einer bestimmten Population; diese sei $5/10^8$ (vgl. Kingston¹⁾).

$(P(D_1/\neg H))$ muß nicht statistisch sein, bzw., wenn statistisch, dann nicht unbedingt $5/10^8$ ²⁾

10. Berechnung:

$$\begin{aligned} (1) P(H_1/D_1) &= \frac{1 \cdot 0.9}{1 \cdot 0.9 + 5 \cdot 10^{-8}} = 0.9 / (0.9 + 5 \cdot 10^{-9}) \\ &= 0.9 / 0.900000005 = 0.999999994. \end{aligned}$$

$$P(H_1/\neg D_1) = 0 \quad (\text{wegen } P(\neg D_1/H_1) = 0);$$

$$P(H_1/P*(D_1)) = 0.999999994 \cdot 0.8 + 0 \cdot 0.2 \approx 0.8.$$

1) Vgl. C.R. Kingston, Probability and Legal Proceedings, in: 57 Journal of Criminal Law, Criminology and Police Science (1966), S.93-98.

2) Die statistische Wahrscheinlichkeit ändert sich außerdem, wenn man berücksichtigt, daß das Haar der Täterin jetzt nicht einfach blond, sondern dunkelblond ist.

(2) $P(H_2/D_2) = 1$; die Begründung kann etwa so lauten: Da die Täterin Mrs. Brooks zu Boden gestoßen hat, ist die Gewalteinwirkung erwiesen und da ich an der Glaubwürdigkeit von Mrs. Brooks keinen Zweifel habe, ist das Datum D_2 seinerseits sicher.

(3) Und schließlich wollen wir die Wahrscheinlichkeit der zusammengesetzten Hypothese H bestimmen.

$$P'(H) = P(H_1 \wedge H_2 // P^*(D_1) \wedge D_2) = \\ P(H_2/H_1 \wedge D_2) \cdot P(H_1 // P^*(D_1))^{1)}$$

Nehmen wir die Unabhängigkeit an, so ist $P'(H) = 0.8$.

Die obige Berechnung zeigt u.a., daß auch dann, wenn die Daten der Sachverhaltshypothese eine sehr hohe Wahrscheinlichkeit verleihen, die letztere nicht größer sein kann als die Zuverlässigkeit der Daten, aufgrund derer die Hypothese beurteilt wird.

Diese Berechnung ist auch insofern interessant, als die Apriori-Wahrscheinlichkeit der Hypothese H_1 im obigen Fall für das Endergebnis ziemlich irrelevant ist und der Wert $P^*(D_1)$ direkt auf das Endergebnis durchschlägt.

1) Berechnung:

$$P'(H) = P(H_1 \wedge H_2 // P^*(D_1) \wedge D_2) = P(H_1 \wedge H_2 / D_1 \wedge D_2) \cdot P^*(D_1) + P(H_1 \wedge H_2 / \neg D_1 \wedge D_2) \cdot \\ P^*(\neg D_1) = P(H_1 / D_1 \wedge D_2) \cdot P(H_2 / H_1 \wedge D_1 \wedge D_2) \cdot P^*(D_1) + P(H_1 / \neg D_1 \wedge D_2) \cdot \\ P(H_2 / H_1 \wedge \neg D_1 \wedge D_2) \cdot P^*(\neg D_1).$$

Wegen $P(H_1 / D_1 \wedge D_2) = P(H_1 / D_1)$, $P(H_2 / H_1 \wedge D_1 \wedge D_2) = P(H_2 / H_1 \wedge D_2)$,

$P(H_1 / \neg D_1 \wedge D_2) = P(H_1 / \neg D_1)$, $P(H_2 / H_1 \wedge \neg D_1 \wedge D_2) = P(H_2 / H_1 \wedge D_2)$ folgt:

$$P'(H) = P(H_1 / D_1) \cdot P^*(D_1) \cdot P(H_2 / H_1 \wedge D_2) + P(H_1 / \neg D_1) \cdot P^*(\neg D_1) \cdot P(H_2 / H_1 \wedge D_2) \\ = P(H_2 / H_1 \wedge D_2) \cdot ((P(H_1 / D_1) \cdot P^*(D_1) + P(H_1 / \neg D_1) \cdot P^*(\neg D_1)) = P(H_2 / H_1 \wedge D_2) \cdot \\ P(H_1 // P^*(D_1)).$$

Ein fiktiver Fall¹⁾

1. Die Sachverhaltshypothese lautet:

Der Angeklagte X hat den Mord (an Y) begangen.

Im vorliegenden Fall handelt es sich um eine zusammengesetzte Sachverhaltshypothese, denn Mord ist ein qualifizierter Tatbestand. Man muß nicht nur nachweisen, daß X Y getötet hat, sondern auch, daß X das z.B. heimtückisch oder grausam getan hat oder um eine andere Tat zu verdecken usw.²⁾

Die einfachen Hypothesen lauten somit:

- 1) X hat Y getötet
- und
- 2) X hat das auf eine durch den Paragraph 211 StGB qualifizierte Weise (z.B. grausam) getan.

2. Das gesamte vorliegende Datum (Schilderung des Falls und anderer Tatsachen, die im Prozeß eine Rolle spielen):

Am Abend des 29. Februar 1987 gegen 21 Uhr fand ein Spaziergänger, namens Collins, die Leiche einer jungen Frau (später erwies es sich, daß es sich bei dieser um Frau Y handelt) im Strassengraben, nicht weit von ihrem Haus entfernt. Er alarmierte sofort die Polizei.

Am Tatort angekommen, entdeckte die Polizei ein Messer mit Blutspuren. Die Untersuchung der Leiche an Ort und Stelle ergab, daß sie völlig angezogen war und keine Spuren aufwies, die auf eine Vergewaltigung bzw. auf einen Vergewaltigungsversuch hindeuteten. (Somit schied ein Vergewaltigungsverbrechen und Mord als Verdeckungstat aus). Die Leiche war jedoch übel zugerichtet; sie wies ca. 20 Messerstiche auf.

Die Untersuchung des Messers ergab erstens, daß es Blutspuren des Opfers aufwies (und daher die mutmaßliche Tatwaffe darstellte), und zweitens, daß ein darauf festgestellter Fingerabdruck dem der rechten Hand des Angeklagten X (des Freundes der Frau) ähnlich war. Es handelte sich allerdings um einen sehr undeutlichen Fingerabdruck (als ob man versucht hätte, ihn abzuwischen); der herangezogene Experte konnte lediglich sagen, daß nur jeder Tausendste einen solchen Fingerabdruck haben könne.

Die Untersuchung der Leiche zeigte außerdem, daß die Frau seit mindestens 5 Stunden und höchstens 1 Tag tot sein konnte.

Der Angeklagte bestritt vor Gericht die Tat heftig, zeigte aber kein Mitleid mit dem Opfer. Er konnte auch ein Alibi anführen für den Tattag. Er sagte, er sei die ganze in Frage kommende Zeit mit einem Freund zusammen gewesen. Der Freund konnte die Behauptung des Angeklagten glaubhaft bestätigen. (Es ist jedoch nicht ausgeschlossen, daß der Angeklagte sich mit dem Freund abgesprochen hat).

¹⁾ Dies ist eine etwas ausführlichere Version des von Finkelstein Fairley zur Einführung des Bayes-Theorems geschilderten Falls; vgl. S. 124f dieser Arbeit.

²⁾ vgl. § 211 StGB.

Ein anderer Zeuge sagte aus, daß der Angeklagte ihm gegenüber nach einem Streit mit dem Opfer, die Tat angekündigt hätte, genauer, er habe ihm gesagt, daß er die Frau hasse und eines Tages umbringen würde.

Ein weiterer Zeuge machte die Aussage, daß er ein ähnliches Messer beim Angeklagten gesehen hätte, war sich jedoch dessen später nicht mehr so sicher.

Auch andere Zeugen bestätigten, daß der Angeklagte sehr oft mit dem Opfer gestritten und es auch geschlagen hat. Der Nachbar der beiden sagte aus, daß am Vorabend der Tat die beiden einen heftigen Streit miteinander gehabt hätten; er konnte sich aber nicht der genauen Uhrzeit des Streits entsinnen.

Der Angeklagte gab den Streit zu, sagte aber, daß er gleich darauf, gegen 19.30 zu seinem Freund gefahren sei, bei dem er gegen 20 Uhr angekommen wäre (die Fahrzeit beträgt 30 min.). Der Freund bestätigte das.

Darauf angesprochen, warum er das Verschwinden der Frau nicht bei der Polizei gemeldet habe, antwortete der Angeklagte, daß diese vorhatte, zu ihrer Mutter aufs Land zu fahren. Die Mutter bestätigte das.

3. Das für die Beurteilung relevante Datum:

1) Ein Fingerabdruck am Messer (der Tatwaffe), der demjenigen des Angeklagten ähnlich ist.

2) Nach Expertenaussage kommt ein solcher Fingerabdruck ein Mal unter tausend Fingerabdrücken vor.

3) Die Aussage eines Zeugen, die besagt, daß der Angeklagte eine solche Tat erwogen hat.

4) Die Aussage desselben Zeugen, die besagt, daß der Angeklagte das Opfer haßte.

5) Die Aussage eines weiteren Zeugen, der sich dessen unsicher ist, ob er ein ähnliches Messer beim Angeklagten gesehen hat oder nicht.

6) Andere Zeugen sagen aus, daß der Angeklagte sehr oft mit dem Opfer gestritten und es auch geschlagen hat.

7) Der Nachbar der beiden sagte aus, daß die beiden am Vorabend des Tattages einen heftigen Streit miteinander hatten; er konnte jedoch keinen genauen Zeitpunkt des Streits nennen.

8) Der Zeitraum, innerhalb dessen die Tat geschehen konnte, liegt zwischen 21 Uhr am 28. Februar und 16 Uhr am 29. Februar.

9) Mord als Verdeckung eines Sexualverbrechens scheidet aus.

10) Tötung in Affekt während des Streits am Vorabend ist nicht ausgeschlossen.

11) Die Leiche ist grausam zugerichtet (20 Messerstiche).

12) Der Angeklagte zeigte kein Mitleid mit dem Opfer.

13) Eine glaubhafte Aussage eines Freundes des Angeklagten, die das Alibi des Angeklagten für den Tag und die Aussage, daß dieser bei ihm am 28. Februar gegen 20 Uhr ankam, bestätigt; es ist jedoch nicht ausgeschlossen, daß der Angeklagte sich mit dem Freund abgesprochen hat.

4. Folgende Daten sind für die Annahme relevant, daß der Fingerabdruck am Messer dem Angeklagten gehört: 1), 2), 5).

Folgende Daten sind für die Hypothese relevant, daß der Angeklagte die Frau getötet hat: zum einen die Information, daß der Fingerab-

1) Dies ist eine explizite Darstellung der Schlußfolgerung aus 7), die für die Beurteilung, ob Mord vorliegt oder nicht, wichtig ist.

druck am Messer dem Angeklagten gehört und zum anderen 1), 2), 3), 4), 5), 6), 7), 8), 9), 13).

Folgende Daten sind für die Annahme relevant, daß es sich um Mord handelt: 10), 11), 12).

Würde man Haß aus psychologischen Erwägungen für den Nachweis des Mordes¹⁾ als relevant annehmen (sei es als "niedriger Beweggrund", sei es im Gegenteil als Entlastung vom Mordvorwurf im Sinn eines Affekts), so wäre hier auch das Datum 4) zu berücksichtigen. Solange das Phänomen Haß (als ein für den Mord relevantes Datum) sich nicht in der vorliegenden Situation eindeutig interpretieren läßt, muß es aufgrund der Forderung (5) aus der Beurteilung ausscheiden. Hingegen nehmen wir hier an, daß Haß für die Tötung eindeutig relevant ist (d.h. wir vertreten hier eine empirisch-psychologische Hypothese, daß man aufgrund des Hasses eher tötet (oder zu töten geneigt ist) als ohne einen solchen. Wir ziehen also das Datum 4) heran, um nachzuweisen, daß X Y getötet hat.

5. Folgende Daten sind unsicher:

- (1) daß der Fingerabdruck dem Angeklagten gehört,
- (2) daß das Messer dem Angeklagten gehört,
- (3) der Zeitpunkt des Streites am Abend vor der Tat,
- (4) der Besuch beim Freund am Tage des Verbrechens.

6. Wir machen die Annahme, daß das vorliegende Datum für die Beurteilung des Falls hinreichend ist.

7. Methode:

Im vorliegenden Fall handelt es sich um eine zusammengesetzte Sachverhaltshypothese. Denn Mord ist, wie bereits erwähnt, ein qualifizierter Tatbestand. Somit ist das Multiplikationsprinzip anzuwenden. Der Nachweis, daß X Y getötet hat, wird mittels der B_{erw} -Formel geführt: Wir beurteilen die Hypothese, daß X Y getötet hat aufgrund der sicheren Information über das Vorliegen eines ähnlichen Fingerabdrucks am Messer, wobei die unsichere Information über die Identität des Fingerabdrucks mitberücksichtigt wird.

8. Wir ziehen für die Aposteriori-Wahrscheinlichkeitsbeurteilung der Hypothese, daß X Y getötet hat, die unsichere Information, daß der Fingerabdruck auf der Tatwaffe der des Angeklagten ist und die sichere Tatsache, daß ein ähnlicher Fingerabdruck am Messer gefunden ist, heran. Die Apriori-Beurteilung wollen wir im Lichte folgender Daten machen: 3), 4), 6), 7) 8), 9), 13).

Wir müssen außerdem nachweisen, daß der Fingerabdruck auf der Tatwaffe dem Angeklagten gehört. Zur Aposteriori-Wahrscheinlichkeitsbeurteilung müssen wir (nach dem vorher gesagten) die Tatsache heranziehen, daß ein ähnlicher Fingerabdruck am Messer gefunden wurde. Für die Apriori-Wahrscheinlichkeit nehmen wir die Aussage eines Zeugen, die besagt, daß ein ähnliches Messer beim Angeklagten gefunden wurde (also 5)). Außerdem wollen wir für die Beurteilung

1) genauer: eines die Tötung als Mord qualifizierenden Merkmals.

auch die Aussage eines Experten heranziehen, daß ein solcher Fingerabdruck einmal unter Tausend vorkommt (also 2)).

Wir müssen außerdem feststellen, ob es sich um einen Mord oder keinen Mord handelt, d.h. ob bestimmte Merkmale vorliegen, die die Tötung von Y als Mord qualifizieren oder nicht. Wir werden das intuitiv im Lichte folgender Tatsachen machen: 10), 11), 12).

9. Zuordnung von Wahrscheinlichkeitswerten:

Wir kürzen die zusammengesetzte Sachverhaltshypothese:

"X hat den Mord an Y begangen" mit H,

die Hypothese:

"X hat Y getötet" mit H_1 ,

die Hypothese:

"Es liegen die den Mord qualifizierenden Merkmale vor" mit H_2 ab.

Der Ereignisraum der zusammengesetzten Sachverhaltshypothese besteht somit aus 4 elementaren Ereignissen bzw. Hypothesen:

$H_1 \& H_2$, $H_1 \& \neg H_2$, $\neg H_1 \& H_2$, $\neg H_1 \& \neg H_2$.

$H_1 \& H_2$ besagt, daß der Angeklagte der Mörder ist,

$H_1 \& \neg H_2$ " , daß der Angeklagte Y getötet hat, allerdings ohne Mörder zu sein,

$\neg H_1 \& H_2$ " , daß es zwar ein Mord ist, daß der Angeklagte jedoch nicht als Täter in Frage kommt,

$\neg H_1 \& \neg H_2$ " , daß weder der Angeklagte der Täter ist noch es sich bei der obigen Tötung um Mord handelt.

Die Spezifizierung des Ereignisraumes ist wichtig, wenn nach Feststellung des Wahrscheinlichkeitswertes für die Sachverhaltshypothese entschieden wird, ob man den Angeklagten freisprechen oder verurteilen soll. So wird man nur im ersten Fall die Verurteilung wegen Mordes verlangen, im zweiten dagegen 'nur' die wegen Totschlags, im dritten und im vierten Fall muß der Angeklagte aber freigesprochen werden. Sind diese vier Ereignisse aber nur mit gewissen Wahrscheinlichkeiten festgestellt, so hängt die Entscheidung auch davon ab, wie schwerwiegend bzw. wünschenswert bestimmte Konsequenzen solcher Entscheidungen unter Unsicherheit (Risiko) sind.

Weitere Abkürzungen:

G dafür, daß der Fingerabdruck auf der Tatwaffe X gehört;

F " , daß der Fingerabdruck auf der Tatwaffe dem von X ähnlich ist.

Wir nehmen folgende Wahrscheinlichkeitswerte an:

$P(G) = 0.1\bar{2}$; die Begründung kann etwa so lauten: Es ist sehr wahrscheinlich, daß man auf dem Messer die Fingerabdrücke des Besitzers findet; daß aber das Messer im vorliegenden Fall dem Angeklagten gehört, ist sehr unwahrscheinlich (denn der Zeuge, der zunächst ausgesagt hat, daß er ein ähnliches Messer bei dem Angeklagten gesehen hat, war sich dessen später doch unsicher; somit kann die

Wahrscheinlichkeit für das Auffinden des Fingerabdrucks des Angeklagten auf dem Messer im Lichte dieser Information nicht höher sein als die Wahrscheinlichkeit für diese Information (und diese ist ziemlich unsicher).

$P(F/G) = 1$; die Begründung kann etwa so lauten: Wenn der Fingerabdruck auf der Tatwaffe X gehört, dann ist der Fingerabdruck mit sich selbst identisch (also auch ähnlich).

$P(F/\neg G) = 0.001$; die Begründung kann etwa so lauten: Wenn der Fingerabdruck nicht X gehört, dann ist die Wahrscheinlichkeit dafür, daß ein ähnlicher Fingerabdruck darauf gefunden wird, gleich der relativen Häufigkeit für das Auftreten solcher Fingerabdrücke in einer Population, denn dann hat man eigentlich keinen Grund, diesen Fingerabdruck gegenüber den anderen Fingerabdrücken auszuzeichnen¹⁾

$P(F/\neg G)$ kann natürlich auch anders ausfallen, wenn es z.B. Gründe gibt, diese Wahrscheinlichkeit nicht mit der relativen Häufigkeit zu identifizieren, z.B., weil man diese Statistik für die Beurteilung des Falls für irrelevant hält, etwa, wenn feststeht, daß dieser Fingerabdruck in einem bestimmten Teil der Bevölkerung, zu dem auch X gehört, häufiger vorkommt).

Um die Wahrscheinlichkeit von $P(H_1/F)$ zu berechnen, nehmen wir folgende Wahrscheinlichkeitswerte an:

$P(H_1) = 0.5$; die Begründung kann etwa so lauten: Einerseits sagten die Zeugen aus, daß X Y haßte und umzubringen vorhatte, daß X mit dem Opfer sehr oft gestritten und es geschlagen hat; außerdem sagte auch der Nachbar, daß die beiden am Vorabend der Tat heftig miteinander stritten; außerdem wurde die Frau nicht Opfer eines Sexualverbrechens. Aber andererseits konnte der Nachbar nicht den genauen Zeitpunkt des Streites nennen und außerdem konnte der Freund des Angeklagten sehr glaubhaft machen, daß X zusammen mit ihm den ganzen Tagtag verbracht hat; darüber hinaus sagte er aus, daß X bei ihm am 28. Februar bereits um 20 Uhr war; die Tatzeit liegt aber zwischen 21 Uhr am 28. Februar und 16 Uhr am 29. Februar; es ist jedoch nicht ausgeschlossen, daß die beiden sich abgesprochen haben. Insgesamt bin ich mir daher unsicher, ob ich im Lichte dieser Information eher an die Hypothese als an ihre Negation glauben soll, d.h. die Negation der Hypothese ($\neg H_1$) kommt mir ebenso wahrscheinlich vor wie die Hypothese (H_1) selbst.

$P(G/H_1) = 0.3$; die Begründung kann etwa so lauten: Wenn jemand die Tat begeht, so versucht er gewöhnlich seine Spuren zu verwischen (es sei denn, er ist dumm oder es kommt etwas dazwischen); wieso sollte X (der auf mich keineswegs einen beschränkten Eindruck macht) sich anders verhalten. Ich will allerdings bei X die Möglichkeit nicht ausschließen, daß er seine Fingerabdrücke hinterlassen hat, falls er die Tat begangen hat, weil ihm z.B. das Verwischen der Fingerabdrücke nicht gelang. (Außerdem bestätigt auch der Experte, der die Möglichkeit erwog, daß der Täter seine Fingerabdrücke absichtlich verwischen wollte, meine Meinung).

1) Dies war übrigens auch die Begründung von Finkelstein&Fairley, vgl. S.124 dieser Arbeit (Finkelstein&Fairley, S.498ff).

$P(G/\neg H_1) = 0.01$; die Begründung kann etwa so lauten: Wenn X nicht der Täter ist, dann ist es sehr unwahrscheinlich, daß man seinen Fingerabdruck auf der Tatwaffe findet. Außerdem liegen keine Informationen vor, die die Annahme eines höheren Wertes im vorliegenden Fall rechtfertigen würden. Die einzige Information, die das Auffinden seines Fingerabdrucks am Messer wahrscheinlich macht, ohne daß der Angeklagte der Täter zu sein braucht, ist, daß er evtl. der Messerbesitzer ist; dieses ist aber seinerseits sehr unwahrscheinlich.

$P(F/\neg G \wedge H_1) = P(F/\neg G \wedge \neg H_1) = P(F/\neg G) = 0.001$; die Begründung kann etwa so lauten: Wenn der Fingerabdruck nicht X gehört, dann ist die Wahrscheinlichkeit dafür, daß ein ähnlicher Fingerabdruck auf der Tatwaffe gefunden wird, davon unabhängig, ob X Y getötet hat oder nicht, d.h. sie ist gleich der relativen Häufigkeit für das Auftreten solcher Fingerabdrücke in einer Population.

10. Berechnung:

$$\begin{aligned}
 (1) \quad P(H_1/F) &= P(H_1/F \wedge G)P(G/F) + P(H_1/F \wedge \neg G)P(\neg G/F) \\
 &= \frac{P(F \wedge G/H_1)P(H_1)}{P(F \wedge G/H_1)P(H_1) + P(F \wedge G/\neg H_1)P(\neg H_1)} \cdot P(G/F) \\
 &+ \frac{P(F \wedge \neg G/H_1)P(H_1)}{P(F \wedge \neg G/H_1)P(H_1) + P(F \wedge \neg G/\neg H_1)P(\neg H_1)} \cdot P(\neg G/F) \\
 &= \frac{P(F/G \wedge H_1)P(G/H_1)P(H_1)}{P(F/G \wedge H_1)P(G/H_1)P(H_1) + P(F/G \wedge \neg H_1)P(G/\neg H_1)P(\neg H_1)} \cdot \\
 &\quad \cdot \frac{P(F/G)P(G)}{P(F/G)P(G) + P(F/\neg G)P(\neg G)} \\
 &+ \frac{P(F/\neg G \wedge H_1)P(\neg G/H_1)P(H_1)}{P(F/\neg G \wedge H_1)P(\neg G/H_1)P(H_1) + P(F/\neg G \wedge \neg H_1)P(\neg G/\neg H_1)P(\neg H_1)} \cdot \\
 &\quad \cdot \frac{P(F/\neg G)P(\neg G)}{P(F/\neg G)P(\neg G) + P(F/G)P(G)} \\
 &= \frac{1 \cdot 0.3 \cdot 0.5}{1 \cdot 0.3 \cdot 0.5 + 1 \cdot 0.01 \cdot 0.5} \cdot \frac{1 \cdot 0.12}{1 \cdot 0.12 + 0.001 \cdot 0.88} \\
 &+ \frac{0.001 \cdot 0.7 \cdot 0.5}{0.001 \cdot 0.7 \cdot 0.5 + 0.001 \cdot 0.99 \cdot 0.5} \cdot \frac{0.001 \cdot 0.88}{0.001 \cdot 0.88 + 1 \cdot 0.12} \\
 &= 0.97 \cdot 0.99 + 0.41 \cdot 0.007 = 0.96.
 \end{aligned}$$

(2) Als nächstes müssen wir bestimmen, ob die Tötung von Y als Mord qualifizierbar ist.

Wir kürzen die Daten 10), 11) und 12) mit D ab.

Es gilt:

$$\begin{aligned} P'(H) &= P(H_1 \wedge H_2 / F \wedge D) = P(H_1 / F \wedge D) \cdot P(H_2 / H_1 \wedge F \wedge D) = \\ &= P(H_1 / F) \cdot P(H_2 / H_1 \wedge D). \end{aligned}$$

Wir nehmen also an, daß H_1 und D füreinander irrelevant sind, d.h. das mangelnde Mitleid des Angeklagten (12)) und die übel zugerichtete Leiche des Opfers (11)) weisen keineswegs ausgerechnet diesen Angeklagten als den Täter aus¹⁾. Das Datum 10) ("Tötung im Affekt...") ist zwar für H_1 relevant, ist aber bereits innerhalb des Datums, im Lichte dessen H_1 beurteilt wurde, berücksichtigt worden; denn 10) ist nichts anderes als das, was 7) ohnehin als eine mögliche Schlußfolgerung impliziert²⁾. Weiter nehmen wir an, daß H_2 und F füreinander irrelevant sind, was trivialerweise gilt.

Wir bestimmen $P(H_2 / H_1 \wedge D)$ intuitiv.

$P(H_2 / H_1 \wedge D) = 0.9$; die Begründung kann etwa so lauten: Wenn X Y getötet hat, dann bestimmt grausam, denn die Leiche ist übel zugerichtet und vor Gericht zeigte X kein Mitleid mit dem Opfer; andererseits ist aber nicht ausgeschlossen, daß er in Affekt gehandelt hat.

Insgesamt erhalten wir:

$$P'(H) = 0.96 \cdot 0.9 = 0.86.$$

Die Wahrscheinlichkeit, daß X den Mord an Y begangen hat, ist somit 0.86.

1) Würde man allerdings eine psychologische Theorie vertreten, welche mangelndes Mitleid allgemein als Indiz für Tötung wertet, dann wären H_1 und 12) nicht voneinander unabhängig. Diese Fußnote und die Bemerkung auf S.225 dieser Arbeit zeigen uns u.a., wie wichtig neben empirischen und rechtspezifischen Informationen auch der sog. psychologische Hintergrund ("background") für die Beurteilung krimineller Taten ist, so daß ein gut geschulter Psychologe bei Strafprozessen unentbehrlich erscheint.

2) Vgl. S.224 dieser Arbeit.

Abschließende Bemerkungen

1. Zunächst sei hier auf ein Problem eingegangen, das vielleicht manchem Leser bereits aufgefallen ist.

Dieses Problem besteht in folgendem:

Wenn dem Individuum bei der Wahl sowohl von Informationen, im Lichte derer es die Apriori-Wahrscheinlichkeitsbeurteilung der Hypothese vornimmt, wie auch von solchen, die es der Aposteriori-Wahrscheinlichkeitsbeurteilung der Hypothese zugrunde legt, uneingeschränkte Freiheit gelassen ist, dann kann es passieren, daß das Individuum verschiedene Aposteriori-Wahrscheinlichkeitswerte ableitet, je nachdem welche Informationen es für die erstere bzw. für die letztere Beurteilung nimmt. Wo liegt der Fehler?

Die Antwort kann nur lauten: In der inkonsistenten Bewertung von Daten seitens des Individuums (Denn die Reihenfolge der Daten, aufgrund derer man die Hypothese beurteilt, darf das Gesamtergebnis der Beurteilung – wegen der Kommutativität von $\wedge$ (vgl. S. 16 dieser Arbeit) – nicht beeinflussen). Das soll näher erläutert werden.

Angenommen, wir haben 5 verschiedene Informationen: D_1, D_2, D_3, D_4, D_5 . Ferner sei angenommen, daß das Individuum im ersten Fall die Apriori-Wahrscheinlichkeitsbewertung der Hypothese H im Lichte der Informationen D_1 und D_2 vornimmt. Anschließend (mittels des Bayes-Theorems) macht es die Aposteriori-Wahrscheinlichkeitsbewertung von H aufgrund der restlichen Informationen D_3, D_4, D_5 . Nehmen wir an, daß seine Apriori-Wahrscheinlichkeit 0.7 und die Aposteriori-Wahrscheinlichkeit 0.9 betragen.

Im zweiten Fall beurteilt das Individuum die gleiche Hypothese aufgrund der gleichen Informationen, nur daß es diesmal die Apriori-Wahrscheinlichkeitsbewertung im Lichte von D_3 und D_5 macht und somit die Aposteriori-Wahrscheinlichkeitsbewertung aufgrund D_1, D_2, D_4 . Nehmen wir an, daß in diesem Fall seine Apriori-Wahrscheinlichkeit 0.5 und seine Aposteriori-Wahrscheinlichkeit 0.7 betragen.

Also hat das Individuum im zweiten Fall an irgendeiner Stelle im Beurteilungsprozeß ein Datum oder mehrere Daten (bzw. die Hypothese aufgrund dieser Daten) anders bewertet als im ersten Fall. Denn ist (wie im ersten Fall) die Apriori-Wahrscheinlichkeit von H im Lichte von D_1 und D_2 0.7 und leitet das Individuum eine Apriori-Wahrscheinlichkeit von H aufgrund von D_3, D_4 und D_5 von 0.9 ab, so müssen im zweiten Fall, falls die Apriori-Wahrscheinlichkeit von H im Lichte von D_3 und D_5 0.5 beträgt, die restlichen Daten D_1, D_2 und D_4 insgesamt der Hypothese soviel Gewicht verleihen, daß die Wahrscheinlichkeit den Wert 0.9 erreicht.

Nun, so exakt kann kein Mensch sein. Es werden immer irgendwelche unkontrollierbaren Faktoren vorhanden sein, die die Beurteilung je nach der Darstellungsweise (und dazu gehört auch die Reihenfolge der Repräsentation von Daten) beeinflussen. Der Fehler bei verschiedenen Beurteilungen aufgrund der gleichen Daten muß trotz alledem sich in einem 'erträglichen' Rahmen halten. Geschieht das nicht, dann ist das ein Hinweis darauf, daß beim Individuum bei verschiedenen Beurteilungen (je nach Reihenfolge der Daten) stillschweigend vielleicht doch ein zu unterschiedliches sog. 'Hintergrundwissen' evoziert wurde.

(Es empfiehlt sich sogar zur Kontrolle, seine eigene Konsistenz zu überprüfen, indem man die Reihenfolge der Daten bei der Beurteilung verändert).

2. Zum Schluß will ich noch einige Gedanken zur Informationsverarbeitung von mehreren Personen im juristischen Prozeß darlegen.

Die subjektivistische Herstellung intersubjektiver Übereinstimmung aufgrund gemeinsamer Erfahrung ist in der Rechtspraxis nicht durchführbar. Denn dieses Prinzip nimmt auf faktische (d.h. objektive) relative Häufigkeitsfeststellungen Bezug,¹⁾ diese letzteren sind aber in der Rechtspraxis nicht gegeben. Wir müssen somit zu einer anderen Methode greifen, wenn wir einen gemeinsamen Wahrscheinlichkeitswert für die Sachverhaltshypothese erhalten wollen, etwa wenn wir diesen der gemeinsamen rationalen Entscheidungsbildung zugrundelegen wollen. Eine sehr einfache Methode besteht darin, daß man den Mittelwert aus den einzelnen Wahrscheinlichkeitswerten bildet. Zur Veranschaulichung:

Angenommen, drei Individuen (wir kürzen sie mit A, B und C ab) leiten aufgrund der gesamten relevanten Information folgende Aposteriori-Wahrscheinlichkeitswerte ab (H steht für die Hypothese, D für das gesamte relevante Datum, P_i für die Wahrscheinlichkeitsbewertung eines Individuums i, $i = A, B, C$; der untere Index i (= A, B, C) zeigt somit, welche Person die Wahrscheinlichkeitsbewertung vornimmt):

$$P_A(H/D) = 0.8, \quad P_B(H/D) = 0.7, \quad P_C(H/D) = 0.9.$$

Daraus erhalten wir den mittleren Wahrscheinlichkeitswert durch Summierung der einzelnen Wahrscheinlichkeiten und Division durch die Anzahl der Personen (wir kürzen den Mittelwert der einzelnen Wahrscheinlichkeiten mit $P(H/D)$ ab).

Der Mittelwert im obigen Beispiel ergibt sich somit als:

$$P(H/D) = 0.8.$$

1) Zu diesem Prinzip vgl. S. 47f dieser Arbeit.

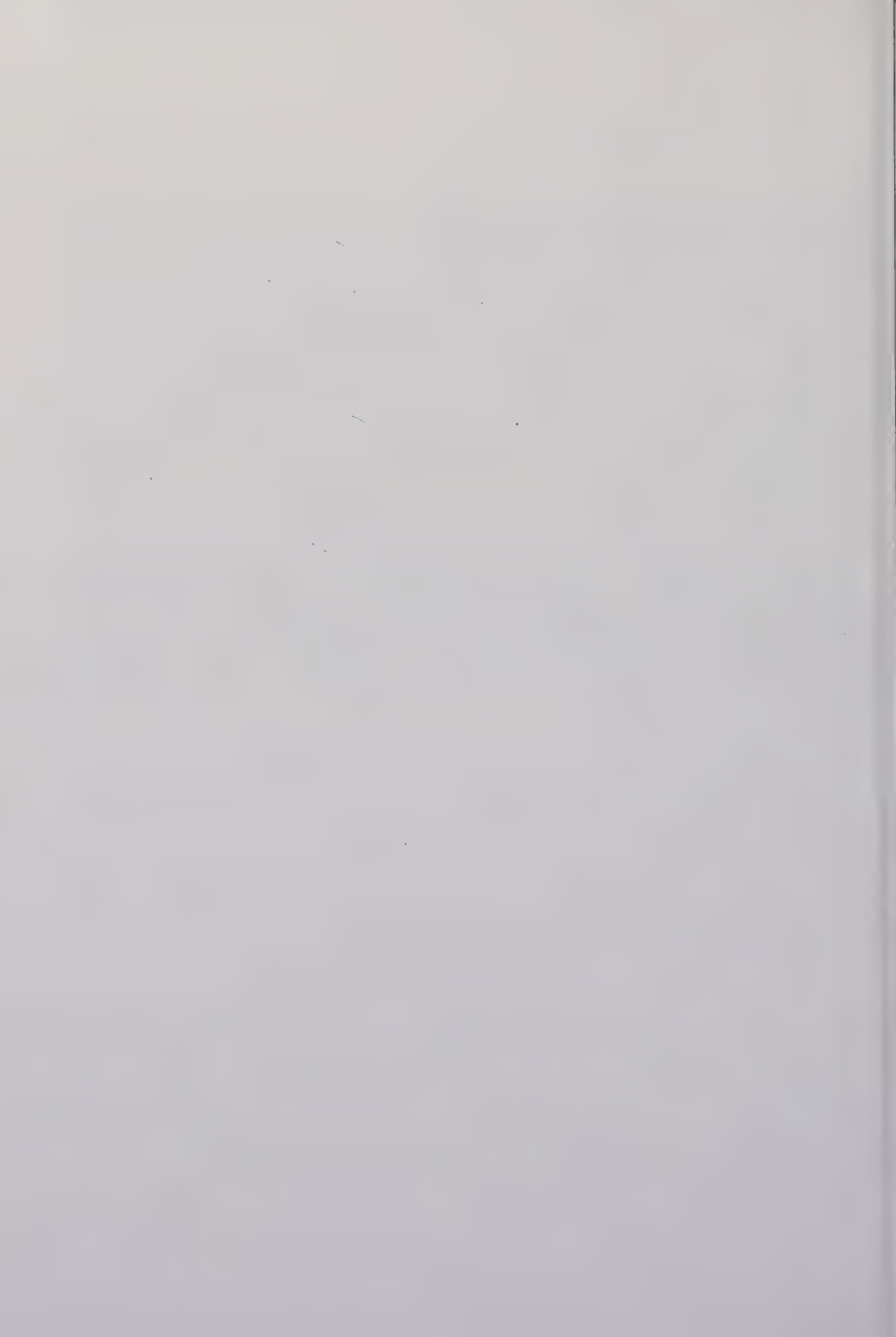
Zusammenfassung

Wie wir gesehen haben, ist die Beurteilung einer juristischen Sachverhaltshypothese aufgrund relevanter Informationen ein subjektiv-probabilistischer Vorgang. In dieser Hinsicht gibt es also keinen Unterschied zwischen einer wahrscheinlichkeitstheoretischen Beurteilung der Sachverhaltshypothese und einer herkömmlichen qualitativen: Denn verschiedene Menschen können sowohl unterschiedlicher Meinung darüber sein, ob ein bestimmter Sachverhalt "mit an Sicherheit grenzender Wahrscheinlichkeit" festgestellt ist, wie auch diesem unterschiedliche (subjektive) Wahrscheinlichkeitswerte zuordnen. Während jedoch die qualitative Beurteilung nicht nur in Bezug auf den Sachverhalt und die für ihn relevanten Tatsachen, sondern auch in Bezug auf die Informationsverarbeitung völlig subjektiv und unkontrollierbar ist, garantiert die wahrscheinlichkeitstheoretische Beurteilung des Sachverhalts eine rationale und kontrollierbare Informationsverarbeitung nach Regeln, die zur konsistenten Verknüpfung subjektiver Annahmen mit anderen Annahmen führen.

Auf den zweiten positiven Aspekt der wahrscheinlichkeitstheoretischen Beurteilung der Sachverhaltshypothese werden wir im II. Teil dieser Arbeit eingehen; wir werden dort sehen, daß die explizite Zuordnung von Wahrscheinlichkeitswerten gerade dann von Bedeutung ist, wenn wir nicht nur die Sachverhaltshypothese präziser beurteilen wollen (als es gewöhnlich der Fall ist), sondern darüber hinaus unter Berücksichtigung der Unsicherheit (Risiko) rational entscheiden wollen.

II. TEIL

RATIONALE JURISTISCHE ENTSCHEIDUNGSTHEORIE



II. Teil: Rationale juristische Entscheidungstheorie

Einleitung

Im I. Teil dieser Arbeit ging es uns u.a. darum, die Besonderheiten der juristischen Argumentationssituation herauszuarbeiten. Wir stellten dabei fest, daß es sich hier um einen Bereich des induktiven (intuitiven, alltäglichen) Rasonierens handelt: Denn es sind v.a. subjektive Meinungen und Einschätzungen, aufgrund derer juristische Urteile gebildet werden. Diese sind in der Regel vage, mehrdeutig und oft nicht einmal nachvollziehbar. Und nicht nur das: Die Verarbeitung solcher subjektiver Meinungen und Einschätzungen zum Gesamturteil bleibt in der herkömmlichen Rechtspraxis völlig im Dunkeln; sie basiert lediglich auf einer völlig verschwommenen 'Disposition', die gewöhnlich als das sog. "Rechtsgefühl des Richters" umschrieben wird.

Demgegenüber wurde hier die Auffassung vertreten, daß juristische Urteile präzisierungsbedürftig sind - will man Korrektheit, Unparteilichkeit und Gerechtigkeit der Rechtspraxis besser gewährleisten -, und - anhand der Wahrscheinlichkeitstheorie - präzisiert werden können. Die Wahrscheinlichkeitstheoretische Beurteilung juristischer Sachverhalte liefert dabei nicht nur die Präzisierung des subjektiven, intuitiven Urteils (bzw. Gefühls) und eine korrekte Verarbeitung einzelner juristischer Tatsachen zum Gesamtbild, sondern ermöglicht auch eine bessere intersubjektive Nachvollziehbarkeit der einzelnen Urteilsbildung.

Die zweite Besonderheit der juristischen Argumentationspraxis besteht - wie bereits im I. Teil erwähnt - darin, daß sie einen Bereich der Entscheidungsbildung darstellt: Es sind letztlich (für die Betroffenen und die Öffentlichkeit wichtige) Entscheidungen, die aufgrund der subjektiven Beurteilung von juristischen Sachverhalten gewonnen werden. Auch diese bzw. allgemeiner: der gesamte Prozeß der juristischen Entscheidungsbildung ist/sind präzisierungsbedürftig (will man, wie oben erwähnt, die Korrektheit, Unparteilichkeit und Gerechtigkeit der Rechtspraxis besser garantieren) und kann/können im Rahmen einer rationalen bzw. normativen Entscheidungstheorie präzisiert werden.

Während wir im I. Teil dieser Arbeit auf den ersten Aspekt der Rationalisierung der juristischen Urteils- und Entscheidungsbildung, d.h. auf die Wahrscheinlichkeitstheoretische Beurteilung der Sachverhaltsfrage, eingegangen sind, wollen wir uns nun mit dem zweiten, nicht minder wichtigen Aspekt befassen, nämlich mit der Bildung rationaler juristischer Entscheidungen. Der Zusammenhang zwischen beiden besteht darin, daß erst die Wahrscheinlichkeitsbeurteilung eines juristischen Sachverhalts eine entscheidungstheoretische (d.h. rationale) Lösung eines juristischen Falls ermöglicht. Andererseits erfährt die (Wahrscheinlichkeitstheoretische) Beurteilung des juristischen Sachverhalts erst in Hinblick auf ihre praktische Auswirkung (richterliche Entscheidung und die Konsequenz dieser Entscheidung) ihren vollen Sinn.

Neben dem Begriff der subjektiven Wahrscheinlichkeit (repräsentiert durch eine Wahrscheinlichkeitsfunktion) ist der Begriff der subjektiven Nützlichkeit bzw. Wünschbarkeit (repräsentiert durch eine Nutzenfunktion) der zweite wichtige Begriff der Entscheidungstheorie. Der Nützlichkeitsbegriff ist insofern wichtig, als in einer Entscheidungssituation der Erfolg unserer Entscheidung oder Handlung nicht nur davon abhängt, daß dieses oder jenes Ereignis eintritt, sondern auch davon, wie wünschenswert uns diese oder jene Konsequenz einer Entscheidung in Verbindung mit verschiedenen Ereignissen erscheint. Das Hauptanliegen dieses Teils der Arbeit besteht somit darin, nicht nur eine rationale juristische Entscheidungstheorie aufzustellen, sondern auch die Kategorie des "juristischen Nutzens" herauszuarbeiten.

Inhaltsüberblick über den II. Teil

In diesem Teil wird eine rationale juristische Entscheidungstheorie aufgestellt. Es handelt sich dabei um ein zweistufiges Entscheidungsmodell. Die erste Stufe betrifft die Wahl einer juristischen Maßnahme (eines Strafrahmens), was eine sog. 'Entscheidungswahl unter Risiko' darstellt, die zweite die Konkretisierung des Strafrahmens, d.h. die Wahl einer bestimmten Strafe innerhalb des Strafrahmens, was eine sog. 'Entscheidungswahl unter Sicherheit' ist. Die Theorie, die sich mit der rationalen Wahl einer juristischen Maßnahme befaßt, wird axiomatisch aufgebaut. Es erweist sich nämlich, daß die Theorie von Savage, die wir der juristischen Entscheidungstheorie zugrunde legen, noch zu weit angelegt ist und ohne Restriktionen zu juristisch unerwünschten Folgen führt. Es werden aus diesem Grund vier Postulate aufgestellt, die die Theorie von Savage in geeigneter Weise einschränken.

Der Theorie der rationalen Strafzumessung, die sich mit der Wahl einer bestimmten Strafe innerhalb des Strafrahmens befaßt, wird das von uns modifizierte Strafzumessungsmodell von Haag zugrundegelegt. Da das Hauptziel der juristischen Strafzumessung in der Bestimmung einer "tatschuldangemessenen" Strafe besteht, wird dies auch unser Hauptanliegen in diesem Bereich der rationalen Entscheidungsbildung sein. Wir werden eine numerische Berechnungsmethode für die Bestimmung der tatschuldangemessenen Strafe vorschlagen.

Im folgenden wollen wir den Inhalt der einzelnen Kapitel dieses Teils kurz skizzieren.

Das erste Kapitel befaßt sich mit den Grundlagen einer rationalen juristischen Entscheidungstheorie. In §1 werden einige intuitive Gedanken zugunsten einer entscheidungstheoretischen Rekonstruktion der Strafrechtspraxis angeführt, die drei Begriffe "Entscheidung unter Sicherheit", "Entscheidung unter Risiko" und "Entscheidung unter Unsicherheit" erläutert und ferner die Leser in die Entscheidungsbildung unter Risiko anhand eines nicht-juristischen Beispiels eingeführt sowie mit dem elementaren Rüstzeug einer solchen Entscheidungsbildung - wie Konsequenzen-, Nützlichkeits- und Wahrscheinlichkeitsmatrix, Präferenzordnung zwischen Handlungen, die Regel von Bayes bekannt gemacht. Außerdem wird auf die Normierung der Nutzenordnung und auf eine graphische Lösung des Problems der Maximierung des Erwartungsnutzens eingegangen.

In §2 untersuchen wir kritisch einige von der juristischen Seite aufgestellten Entscheidungsmodelle. Es stellt sich heraus, daß diese Modelle alle an erheblichen Mängeln leiden: entweder sind sie nicht konsistent oder nicht adäquat bzw. nicht ausreichend. Anhand dieser Untersuchung werden wir außerdem (vorläufig informell) herausstellen, welchen Anforderungen eine adäquate und konsistente juristische Entscheidungstheorie genügen muß.

In §3 befassen wir uns tiefer mit der Frage, ob man eine Theorie der rationalen juristischen Entscheidungsbildung im Rahmen einer bestimmten, von der Entscheidungstheorie vorgeschlagenen Konzeption entwickeln kann. Es werden dabei drei Konzeptionen untersucht: diejenige von Savage, diejenige von Fishburn und diejenige von Jeffrey. Es erweist sich, daß nur die Konzeption von Savage für unsere Zwecke in Frage kommt. Allerdings ist das Entscheidungsmodell von Savage

noch zu weit: So wird darin u.a. keineswegs verhindert, daß eine juristische Entscheidung (unerwünschterweise) entsprechend dem unwahrscheinlichsten strafbaren Sachverhalt als die beste oder auch eine (nicht minder juristisch unerwünschte) Indifferenz zwischen den Maßnahmen abgeleitet wird.

Im zweiten Kapitel wird eine rationale Theorie juristischer Einzelentscheidungen entwickelt. In §1 werden die im ersten Kapitel eingeführten entscheidungstheoretischen Begriffe auf eine juristische Entscheidungssituation angewandt. Es geht uns dabei v.a. darum, die entscheidungstheoretischen Grundentitäten als juristische Begriffe zu begreifen und als solche zu rekonstruieren. Außerdem wird hier die Problematik der sog. 'qualifizierten Tatbestände' erörtert. Um auch von solchen - und nicht nur von den Grundtatbeständen - im Modell sinnvoll reden zu können, werden die letzteren mit sog. "elementaren Sachverhalten" und die ersteren mit "Ereignissen" identifiziert. Darüber hinaus machen wir von einer formalen Konstruktion Gebrauch, um eine Entscheidungssituation in den Griff zu bekommen, in welcher neben den Grundtatbeständen gleichzeitig auch qualifizierte Tatbestände in Betracht gezogen werden.

In §2 wollen wir den zentralen Begriff des "juristischen Nutzens" klären. Dieser wird unter dem sozialen Aspekt gesehen, unter Berücksichtigung sowohl der Belange der durch die Straftat Betroffenen wie auch derer der Gesellschaft und auch derer des Angeklagten. Um der Theorie des juristischen Nutzens eine einigermaßen gesicherte Grundlage zu bieten, knüpfen wir dabei an die in der Rechtspraxis vorhandenen strafrechtlichen Theorien von Sinn und Zweck der Strafe an. Allerdings wird der Grundrahmen dieser Theorie eine (in der Rechtspraxis in dieser Form noch nicht formulierte) Gerechtigkeitsauffassung sein, die davon ausgeht, daß eine Straftat eine (negative) Veränderung des sozialen Gleichgewichts oder des juristischen Status Quo ist. Dementsprechend wird die (gerechte) Strafe als die formale Wiederherstellung dieses Status Quo verstanden.

Alle übrigen juristischen Ziele wie Verbrechensbekämpfung und -verhütung, Resozialisierung, Vergeltung, Genugtuung des Geschädigten, die Befriedigung des Sicherheitsbedürfnisses der Gesellschaft (etwa durch die Abschreckung potentieller Täter) usw. werden somit auf das fundamentale Ziel "Gerechtigkeit" relativiert. Wir werden uns jedoch nicht mit der juristischen (inhaltlichen) Interpretation all dieser Ziele befassen, sondern rein formal verschiedene Möglichkeiten untersuchen, diese durch entsprechende Nützlichkeitsmatrizen zu repräsentieren. Auf diese Weise können wir auch besser bestimmen, in welcher Form derartige Vorstellungen Einfluß auf die Entscheidungsbildung ausüben. Wir werden sehen, daß nur Gerechtigkeit als Ziel den juristischen Grundsatz der Gleichbehandlung vor dem Gesetz formal (wegen der Eindeutigkeit der "gerechten" Nutzenzuordnung) garantiert.

§3 ist insofern zentral für die ganze Arbeit, als hier einige Forderungen zusätzlich zu den entscheidungstheoretischen Postulaten aufgestellt und begründet werden, die aus dem entscheidungstheoretischen Grundmodell (Savage-Modell) eine juristische Entscheidungstheorie entstehen lassen. Insgesamt werden hier vier Postulate formuliert, die die rationale Entscheidungsbildung juristisch einschränken sollen. Bei diesen vier Postulaten handelt es sich um sog. minimale Forderungen, die zwar bestimmte, aus dem Grundmodell

sich ergebende unbefriedigende Folgen verhindern und neben dem objektiven auch den subjektiven Sachverhalt in die rationale Entscheidungsbildung einbeziehen, gewisse juristische Forderungen aber, wie etwa die, daß die Entscheidung nur dann ableitbar sein soll, wenn der entsprechende Sachverhalt mit "an Sicherheit grenzender Wahrscheinlichkeit" festgestellt ist, nicht erfüllen (Es wird gezeigt, daß die letztere Forderung sogar formal zu unsinnigen Einschränkungen bzgl. der Nützlichkeitsmatrix führt).

Während sich die ersten zwei Kapitel ausschließlich mit der Problematik der rationalen Entscheidungsbildung einer einzelnen Person (eines einzelnen Richters) befassen, werden in den letzten beiden zwei andere Aspekte des juristischen Entscheidungsprozesses angesprochen.

Im dritten Kapitel machen wir uns einige Gedanken über die kollektive juristische Entscheidungsbildung. Wir beschränken uns dabei von vornherein auf sog. Konfliktsituationen, in denen Meinungsverschiedenheiten bzgl. der juristischen Zielsetzung auftreten. Es wird allerdings keine Theorie kollektiver juristischer Entscheidungen entwickelt, sondern wir diskutieren ganz kurz verschiedene Lösungsansätze juristischer Konflikte. Alle diese Lösungsansätze laufen darauf hinaus, juristische kollektive Entscheidungen im Sinn nicht-kompetitiver kooperativer Mehrheitsentscheidungen zu verstehen, d.h. will ein Richter an der gemeinsamen Entscheidungsbildung teilnehmen, so muß er einen nicht-kompetitiven, kooperativen Standpunkt innerhalb des Richterorgans beziehen (Es sei hier noch angemerkt, daß dieses Kapitel wegen des Umfangs der Arbeit zu kurz gekommen ist und daß in diesem Bereich (kollektiver juristischer Entscheidungen) noch viel zu tun ist).

Im vierten Kapitel schließlich befassen wir uns mit der Strafzumessung, d.h. mit der Bestimmung der konkreten Strafe innerhalb des gesetzlichen Strafrahmens im Einzelfall. Wie bereits oben erwähnt, handelt es sich dabei um eine 'Entscheidungswahl unter Sicherheit'. Im Mittelpunkt des von uns aufgestellten Strafzumessungsmodells (Optimierungsmodells) steht die Bestimmung der tatschuldangemessenen Strafe innerhalb des Strafrahmens. Wir verknüpfen das Optimierungsmodell mit dem in den ersten beiden Kapiteln entwickelten entscheidungstheoretischen Modell der Wahl einer Maßnahme (eines Strafrahmens) unter Risiko, indem wir einerseits die Menge aller möglichen konkretisierten Maßnahmen mit dem Strafrahmen identifizieren und andererseits Wahrscheinlichkeiten für strafbare Tatbestände (Risiko-Modell) als Gewichte bei der Berechnung einer konkreten Strafe (Optimierungsmodell) berücksichtigen.

Die Tatsache, daß vielen Juristen die Zuordnung von numerischen Werten zu subjektiven Urteilen unmöglich erscheint, trägt dazu bei, warum formale bzw. mathematische Methoden abgelehnt werden und man in der Rechtspraxis noch immer intuitiv-gefühlsmäßig vorgeht. Dieser Irrtum läßt sich korrigieren, indem man zeigt, daß auch unsere subjektiven Überlegungen quantifizierbar sind. Im Anhang geben wir deswegen verschiedene Verfahren an, die sich mit der Metrisierung subjektiver Wahrscheinlichkeitsurteile und subjektiver Nützlichkeitsurteilungen befassen. Allerdings verlangt das von einem Juristen die Einarbeitung in dieses Gebiet. Außerdem werden im Anhang die von Neumann-Morgenstern-Axiome und diejenige von Savage angegeben und erklärt.

1. Kapitel. Grundlagen einer rationalen juristischen Entscheidungstheorie

§1. Einführung in die rationale Entscheidungstheorie

Wir wollen hier einige intuitive Gedanken zugunsten der Anwendung der rationalen Entscheidungstheorie auf die Strafrechtspraxis formulieren und ferner die drei wichtigsten Begriffe "Entscheidung unter Risiko", "Entscheidung unter Sicherheit" und "Entscheidung unter Unsicherheit" erläutern. Außerdem wird der Leser hier in die Entscheidungsbildung anhand eines nicht-juristischen Beispiels eingeführt und mit den elementaren Grundlagen und Begriffen dieser Entscheidungsbildung - wie Konsequenzen-, Nützlichkeits- und Wahrscheinlichkeitsmatrix, Präferenzordnung zwischen Handlungen, die Regel von Bayes - bekannt gemacht. Abschließend sollen die Normierung der Nutzenordnung und eine graphische Lösung des Problems der Maximierung des Erwartungsnutzen an einem Beispiel kurz erläutert werden.

Wir beginnen mit der Annahme, daß es sich bei einem Richter um eine Person handelt, die bei ihrer Entscheidung möglichst rational vorgehen will. Kurz, ein Richter ist in unseren Vorstellungen ein rational Handelnder. Ob diese Annahme stimmt, d.h. ob das tatsächliche Verhalten dem entspricht, was in dieser Situation als rationales Verhalten angemessen gewesen wäre, soll hier nicht diskutiert werden. Die Beantwortung dieser Frage gehört nicht mehr in den Bereich der normativen (rationalen), sondern in den der sog. deskriptiven Entscheidungstheorie¹⁾

Wir akzeptieren also die Hypothese von der Form "der Richter X ist ein rational Handelnder", ohne auf ihre Berechtigung weiter einzugehen.

Von einem rational Handelnden erwartet man gewöhnlich, daß er die Entscheidungssituation, d.h. alle seiner Meinung nach möglichen Umstände und Handlungen und die daraus resultierenden Konsequenzen, übersieht (Die Entscheidungssituation kann natürlich so komplex

1) Der Unterschied zwischen der normativen (rationalen) und der deskriptiven Entscheidungstheorie besteht darin, daß erstere gewisse Regeln (Rationalitätskriterien) für das konsistente und konsequente Verhalten formuliert, während letztere 'nur' das tatsächliche, empirisch-feststellbare Verhalten eines Individuums beschreibt. Wie bereits gesagt, geht es in dieser Arbeit nicht um die Legitimation der tatsächlichen juristischen Vorgehensweise, sondern um deren Rationalisierung. Dies erklärt auch unser jetziges Hauptanliegen, das in der Aufstellung einer normativen bzw. rationalen juristischen Entscheidungstheorie besteht (Im folgenden werden wir - soweit es möglich ist - nur den Terminus "rationale" juristische Entscheidungstheorie gebrauchen, um eine Verwechslung mit der sog. "normativen juristischen Begründungstheorie", die sich mit der Auslegung von juristischen Gesetzen und Normen befaßt, zu vermeiden).

sein, daß der Handelnde nicht ohne weiteres imstande sein wird, alle möglichen Umstände, Handlungen und Konsequenzen anzugeben. Da es sich jedoch nicht um eine objektive Entscheidungssituation, sondern um eine solche im Bewußtsein der rationalen Person handelt, d.h. "alle möglichen Umstände..." sind nicht alle objektiv möglichen, sondern nur solche, die dem Handelnden bekannt und seiner Meinung nach möglich und erschöpfend sind, ist die Sachlage weniger problematisch, als es zunächst erscheinen mag).

In der Entscheidungstheorie wird üblicherweise zwischen drei Arten von Entscheidungssituationen unterschieden: Entscheidungssituationen unter Sicherheit, Entscheidungssituationen unter Risiko und Entscheidungssituationen unter Unsicherheit¹⁾ Unter einer Entscheidungssituation unter Sicherheit versteht man eine solche, in der sowohl Umstände (in der Regel ein einziger) als auch Handlungen (Extremfall: eine einzige) nach der Beurteilung des Handelnden mit Bestimmtheit zu gewissen Resultaten führen. In der herkömmlichen Rechtspraxis ist eine solche Entscheidungssituation in Form des sog. 'juristischen Syllogismus' gegeben: Tatbestand ($\hat{=}$ Umstand) und daran anknüpfende Rechtsfolge bzw. Gesetzesvorschrift ($\hat{=}$ Entscheidung, Handlung) haben als Konsequenz die Verurteilung entsprechend dieser Gesetzesvorschrift. Ein weiterer Fall liegt bei der Strafzumessung, die in der Wahl einer konkreten Strafe innerhalb eines vorgegebenen Strafrahmens besteht, vor.

Eine Entscheidungssituation unter Risiko²⁾ ist dadurch charakterisiert, daß jede der in Frage kommenden Handlungen zu verschiedenen möglichen Resultaten führen kann; der Handelnde vermag dabei zwar die für ihn relevanten Konsequenzen seines Handelns nicht genau vorausszusehen, ist jedoch in der Lage, den möglichen Umständen sowie den Folgen seiner Handlungen gewisse subjektive Wahrscheinlichkeiten zuzuordnen.

1) Es ist allerdings hier anzumerken, daß Luce & Raiffa, auf die die Unterscheidung Sicherheit/Unsicherheit/Risiko zurückgeführt werden kann, von einem objektiven Wahrscheinlichkeitsbegriff ausgegangen sind. Für die Subjektivistin, die die Entscheidbarkeit von (subjektiven) Wahrscheinlichkeiten annehmen (vgl. S. 45 dieser Arbeit), ergibt die Situation unter Unsicherheit wenig Sinn.

2) Gelegentlich auch "Entscheidungssituation unter Unsicherheit im weiten Sinne" genannt.

Kann der Handelnde den möglichen Umständen keine Wahrscheinlichkeitswerte zuschreiben, so liegt eine Entscheidungssituation unter Unsicherheit¹⁾ vor. In einer solchen Situation sind dem Handelnden sowohl die möglichen Umstände und in Frage kommenden Handlungen als auch die Konsequenzen (mit deren Wünschbarkeiten) bekannt; dagegen kann er keine Wahrscheinlichkeitsverteilung für die möglichen Umstände (bzw. Resultate) angeben; d.h. der Handelnde weiß nicht, ob der Umstand diese oder jene Wahrscheinlichkeit hat (oder in der qualitativen Sprechweise, ob der Umstand wahrscheinlich oder unwahrscheinlich ist).

Akzeptiert man, daß es sich bei einem juristischen Sachverhaltsurteil um ein subjektives Wahrscheinlichkeitsurteil handelt, und daß außerdem die (wahrscheinlichkeitstheoretische) Beurteilung des juristischen Sachverhalts erst in Hinblick auf ihre praktische Auswirkung (Entscheidung und Konsequenz der Entscheidung) ihren vollen Sinn erfährt²⁾, so muß man auch akzeptieren, daß eine juristische Entscheidungssituation eine solche mit mehreren Entscheidungsalternativen und mehreren wahrscheinlichen Resultaten - in der neuen Terminologie also eine Entscheidungssituation unter Risiko - ist³⁾.

Gegen die Annahme, daß die Rechtspraxis eine solche Entscheidungssituation unter Unsicherheit sein könnte, spricht der Umstand, daß der Richter dann konsequenterweise seine Entscheidung ohne jeglichen Bezug auf den Sachverhalt fällen müßte (denn dieser wäre dann nach Voraussetzung völlig unbestimmt, d.h. nicht einmal subjektiv-probabilistisch bestimmbar). Dies widerspricht aber der juristischen Wirklichkeit, in der der Richter seine Entscheidung stets an die Beurteilung des Sachverhalts anknüpft; auch wenn der Richter den Sachverhalt nicht verifizieren kann, so vermag er in der Regel anzugeben, ob seiner Meinung nach der Sachverhalt

1) Gelegentlich auch "Entscheidungssituation unter Unsicherheit im engeren Sinne" genannt.

2) Vgl. S.61f, 133 Fn.1, 195f, 233 dieser Arbeit.

3) Für die wahrscheinlichkeitstheoretische Beurteilung einer juristischen Sachverhaltshypothese glauben wir genug Gründe angeführt zu haben. Daß die richterliche Tätigkeit praktisch auf eine Entscheidung ausgerichtet ist, kann auch dann nicht bestritten werden, wenn man die richterliche Tätigkeit hauptsächlich als das Auslegen von Gesetzen begreift. Daraus ergibt sich übrigens, daß man die juristische Entscheidungssituation auf jeden Fall entscheidungstheoretisch rekonstruieren muß.

dafür ausreicht bzw. nicht ausreicht, eine bestimmte Rechtsfolge (in der herkömmlichen Entscheidungssituation) 'abzuleiten'. D.h. der Richter geht bei seiner Entscheidungsbildung, der die Beurteilung des Sachverhalts vorangegangen ist, - auch wenn ihm das nicht voll bewußt ist - von einer probabilistischen Vorstellung hinsichtlich des Bestehens bzw. Nicht-Bestehens des Sachverhalts aus. Die juristische Entscheidungssituation ist somit nicht eine unter Unsicherheit, sondern eine unter Risiko!)

Einer rationalen Entscheidung (Handlung) liegen immer bestimmte Ziel- und Normenvorstellungen zugrunde, die man gewöhnlich als 'Gesamtziel des Handelnden' bezeichnet. In der Entscheidungstheorie gilt es als selbstverständlich, daß in dieses Gesamtziel soziale, moralische und religiöse Normen eingehen. Was die juristische Entscheidungssituation betrifft, so ist hier das Gesamtziel auf wenige durch die Rechtspraxis festgelegte Zielvorstellungen, wie Gerechtigkeit, Verbrechensbekämpfung und -Verhütung, Resozialisierung und dergl., beschränkt. Diesen Zielvorstellungen liegen ihrerseits (un)bestimmte, durch die Rechtspraxis nicht mehr kontrollierbare moralische, soziale und religiöse Auffassungen zugrunde. Wir werden in diesem Kapitel die Ziel- und Normenvorstellungen des rationalen Richters als vorgegebene Daten betrachten, später aber (2.Kap.) auf verschiedene konkrete diesbezügliche Vorstellungen näher eingehen und entsprechend auf die Zuordnung unterschiedlicher Wünschbarkeitswerte zu den Konsequenzen.

Die Entscheidungssituation unter Risiko soll hier an einem nicht-juristischen Beispiel skizziert werden. Um Entscheidungen unter Risiko zu präzisieren, müssen wir von drei Matrizen Gebrauch machen: der Konsequenzen-, Nützlichkeits- (Wünschbarkeits-) und Wahrscheinlichkeitsmatrix. Das nachfolgende Beispiel ist bei Stegmüller²⁾ zu finden.

1) Wir können allerdings auch im juristischen Bereich Entscheidungssituationen unter Unsicherheit vorfinden. Ein Beispiel dafür wäre das Auslegen von Gesetzen.

2) Siehe Stegmüller, Bd.IV, Entscheidungslogik, S.288f. Wir geben den Wortlaut des Beispiels mit geringfügigen Modifikationen wieder.

Die Konsequenzenmatrix:

Angenommen, eine Person X möchte vom Ort y zum Ort z gelangen. Es stehen ihr zwei Möglichkeiten offen: Sie kann entweder mit der Bahn fahren oder ein Flugzeug nehmen. Dies sind die beiden möglichen Handlungen, zwischen denen sie wählen muß. Falls sie sich für die Bahn entschließt, beträgt die Fahrzeit 7 Stunden. Falls sie jedoch ein Flugzeug nimmt, bestehen zwei Möglichkeiten: Wenn es am Zielort nicht neblig ist, so beträgt die Flugzeit 2 Std.. Sollte jedoch der Zielort eingenebelt sein, dann muß das Flugzeug umdirigiert werden, so daß die Fahrzeit insgesamt 16 Std. beträgt. Die folgende Tabelle liefert einen Überblick über sämtliche Möglichkeiten:

Handlungen/Umstände	Zielort nebelfrei	Zielort eingenebelt
Eisenbahn	7 Std.	7 Std.
Flugzeug	2 Std.	16 Std.

Jeder der möglichen Handlungen, zwischen denen X zu wählen hat, entspricht in dieser Tabelle eine Zeile. Jedem der beiden relevanten Umstände (Zielort nebelfrei bzw. Zielort eingenebelt) entspricht eine Spalte. Am Kreuzungspunkt einer Zeile und einer Spalte ist die Konsequenz oder Folge eingetragen, die sich für X ergibt, wenn sie die dieser Zeile entsprechende Handlung wählt und der der Spalte entsprechende Umstand realisiert ist. Im vorliegenden Beispiel bestehen diese Konsequenzen aus Fahrzeiten, gemessen in Stunden vom Zeitpunkt der Abfahrt bzw. des Abflugs vom Ort y. Aus dieser Matrix erhält man die Matrix der Folgen oder die 'Konsequenzenmatrix', wenn man die Zeilen- und Spaltenüberschriften wegläßt und nur die in Zahlen ausgedrückten Folgen einträgt. In unserem Fall sieht die Konsequenzenmatrix so aus:

7	7
2	16

Aus diesem Beispiel können wir das allgemeine Schema abstrahieren. Eine Person möge in einer bestimmten Situation zwischen m Handlungen $h_1, \dots, h_m$ wählen können. Für die für sie interessanten möglichen Konsequenzen (oder Resultate) seien n mögliche 'Weltzustände', kurz Umstände $U_1, \dots, U_n$ genannt, maßgebend. Es sei dabei vorausgesetzt, daß die Person überzeugt ist, daß die n Umstände eine erschöpfende Disjunktion aller möglichen Umstände und die m Handlungen eine erschöpfende Disjunktion aller möglichen Handlungen bilden. Wie die folgende Tabelle zeigt, entspricht einer Handlung h_i und einem Umstand U_k ein Resultat R_{ik} (bzw. eine Konsequenz, in der früheren Sprechweise), wobei der erste untere Index (Zeilenindex) am "R" über die in bestimmter Reihenfolge numerierten möglichen Handlungen läuft und der zweite untere Index (Spaltenindex) über die in bestimmter Reihenfolge numerierten möglichen Umstände. Aus

dieser Tabelle gewinnt man die Konsequenzenmatrix, kurz Matrix $[R_{ik}]$ genannt).

Es sei noch einmal betont, daß es hier nicht auf den objektiven Sachverhalt ankommt, sondern allein auf die subjektiven Überzeugungen des Handelnden, d.h. in der obigen Darstellung schildert die Konsequenzenmatrix (R_{ik}) nicht die objektive Situation, sondern die Situation in der Interpretation des Handelnden.

	U_1	U_2	...	U_n
h_1	R_{11}	R_{12}	...	R_{1n}
h_2	R_{21}	R_{22}	...	R_{2n}
.	.	.		.
.	.	.		.
.	.	.		.
h_m	R_{m1}	R_{m2}	...	R_{mn}

Bei der Aufstellung der obigen Konsequenzenmatrix gingen wir davon aus, daß jedes Resultat R_{ik} eindeutig bestimmt ist, wenn die Handlung h_i und der Umstand U_k festliegen. Mathematisch können wir eine Funktion ϕ annehmen, die jeder Handlung h_i und jedem Umstand U_k das Resultat R_{ik} als Funktionswert zuordnet, d.h. es gilt $R_{ik} = \phi(h_i, U_k)$

Die Nützlichkeitsmatrix oder Wünschbarkeitsmatrix:

Wir setzen voraus, daß X alle möglichen Resultate R_{ik} subjektiv bewertet, d.h. X ordnet jedem der $m \times n$ möglichen Resultate $R_{11}, \dots, R_{mn}$ einen 'subjektiven Wert' oder 'Nutzen' zu. Um die Schilderung zu vereinfachen, nehmen wir weiter ohne Beweis an, der subjektive Wert oder Nutzen könne durch reelle Zahlen charakterisiert werden. Setzt man für jedes mögliche Resultat dessen subjektiven Nutzen ein, so gewinnt man die Nützlichkeitsmatrix (wobei wir die subjektive Nützlichkeit nicht nur als positiven Nutzen durch positive Zahlenwerte charakterisieren, sondern auch als negativen Nutzen durch negative Zahlenwerte darstellen können).

Sei X im obigen Beispiel ein stark beanspruchter Geschäftsmann, für den Zeit Geld bedeutet, d.h. verlorene Zeit ist für ihn verlorenes Geld. Da er beim Reisen die angegebenen Reisezeiten als Verlustzeiten verbuchen muß, liegt es daher nahe, die mit negativen Vorzeichen genommenen Zeiten als 'subjektive Nützlichkeiten' einzutragen¹⁾

1) Vgl. Stegmüller, S.292.

Die Nützlichkeitsmatrix lautet somit:

$$\begin{bmatrix} -7 & -7 \\ -2 & -16 \end{bmatrix}.$$

Um dieses Beispiel zu verallgemeinern, müssen wir annehmen, daß eine von X abhängige Funktion u , genannt Nutzenfunktion, existiert. Die Argumente dieser Funktion sind die möglichen Resultate R_{ik} , die Funktionswerte sind die reellen Zahlen, welche die subjektiven Werte dieser Resultate repräsentieren. Wendet man die Funktion u auf sämtliche Resultate an, die in der Konsequenzenmatrix vorkommen, so erhalten wir die Nützlichkeitsmatrix, abgekürzt $[u(R_{ik})]$ bzw. $[u_{ik}]$. Verallgemeinert sieht somit die Nützlichkeitsmatrix folgendermaßen aus:

$$\begin{bmatrix} u_{11} & u_{12} & \dots & u_{1n} \\ u_{21} & u_{22} & \dots & u_{2n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ u_{m1} & u_{m2} & \dots & u_{mn} \end{bmatrix}$$

wobei die Anzahl der möglichen Umstände n und m die Anzahl der möglichen Handlungen ist

Die Werte der Funktion u für R_{ik} (auch 'subjektive Nützlichkeiten' oder einfach 'Nutzen' von R_{ik} für X genannt), also $u(R_{11}), \dots, u(R_{mn})$, kann man als 'numerische Nutzenordnung' (der Resultate bzw. der Konsequenzen) bezeichnen¹⁾

Die Wahrscheinlichkeitsmatrix

Wir setzen voraus, daß unser Subjekt die Wahrscheinlichkeit des Eintretens jedes einzelnen der n Umstände zu beurteilen vermag. Formal geschrieben kann also X für jeden Umstand U_i die Wahrscheinlichkeit seines Eintretens $P(U_i)$, $i = 1, \dots, n$, angeben. Wie schon mehrmals erwähnt, kommt es hier nicht auf die 'objektiven Wahrscheinlichkeiten' an, sondern allein auf die subjektiven Bewertungen, um zu einer Entscheidung bzgl. der Wahl einer möglichen Handlung zu gelangen.

1) Vgl. Stegmüller, S. 292.

Man unterscheidet zwischen der 'handlungsunabhängigen Wahrscheinlichkeitsmatrix' und der 'handlungsabhängigen Wahrscheinlichkeitsmatrix'. Bei der handlungsunabhängigen Wahrscheinlichkeitsmatrix ist die Wahrscheinlichkeit für das Eintreten eines Umstandes U_i von der Wahl der Handlung unabhängig. Man gibt in diesem Fall nur die Wahrscheinlichkeitsverteilung für die Umstände, also die n Werte $P(U_1) = p_1, P(U_2) = p_2, \dots, P(U_n) = p_n$ an. Die Wahrscheinlichkeitsmatrix sieht dann so aus:

allgemeiner Fall:

$$\begin{bmatrix} p_1 & p_2 & \dots & p_n \\ \vdots & \vdots & & \vdots \\ p_1 & p_2 & \dots & p_n \end{bmatrix} \quad \begin{matrix} \\ \\ m \text{ Zeilen} \end{matrix}$$

n Spalten

in unserem Beispiel:

$$\begin{bmatrix} p & 1-p \\ p & 1-p \end{bmatrix}$$

p ist die Wahrscheinlichkeit, daß der Zielort neubefrei ist;
 $1-p$ ist die Wahrscheinlichkeit, daß er eingeebelt ist.

Ist $p = 9/14$, so lautet die Matrix:

$$\begin{bmatrix} 9/14 & 5/14 \\ 9/14 & 5/14 \end{bmatrix}$$

Im Fall der handlungsabhängigen Wahrscheinlichkeitsmatrix genügt es nicht, nur die Wahrscheinlichkeitsverteilung für die n Umstände zu kennen, sondern es kommt auch auf die gewählte Handlung an. Ein Beispiel zur Illustration (dieses Beispiel ist sowohl bei Jeffrey wie auch bei Stegmüller zu finden¹⁾):

Ein fünfunddreißigjähriger Amerikaner X habe bisher pro Tag mindestens zwei Schachteln Zigaretten geraucht. Er liest eine Statistik der American Cancer Society und wird schwankend. In der Statistik wurde u.a. über die Chancen 35 jähriger Männer berichtet, älter zu werden als 65, je nachdem, ob sie Nichtraucher, Pfeifen- und Zigarrenraucher, schwache, starke oder sehr starke Zigarettenraucher sind. X muß feststellen, daß er nicht die Willenskraft hat, weniger als zwei Schachteln Zigaretten täglich zu rauchen, wenn er überhaupt beim Zigarettenrauchen bleibt. Die einzige Möglichkeit erblickt er darin, evtl. zum Zigarren- und Pfeifenrauchen überzugehen, welches ihm allerdings weniger Genuß bereitet. Er stellt sich aus der erwähnten Statistik die folgende Wahrscheinlichkeitsmatrix zusammen. (Da er an dem Bericht nicht zweifelt, fallen diesmal die subjektiven Wahrscheinlichkeiten mit den in der Statistik erwähnten objektiven relativen Häufigkeiten zusammen):

¹⁾ R. Jeffrey, Logik der Entscheidungen. Wien-München (1967), S. 21f und S. 46f. W. Stegmüller, S. 294. Der Wortlaut des Beispiels ist von Stegmüller mit einer geringfügigen Modifikation übernommen.

	vor dem 65. Lebensjahr sterben	mindestens 65 Jahre alt werden
mindestens zwei Schachteln Zigaret- ten pro Tag rauchen	0.41	0.59
nur Pfeifen und Zi- garren rauchen	0.25	0.75

Das allgemeine Schema sieht diesmal so aus:

$$\begin{bmatrix} p_{11} & \dots & p_{1n} \\ \vdots & & \vdots \\ p_{m1} & \dots & p_{mn} \end{bmatrix}$$

In diesem Fall übertragen sich die Wahrscheinlichkeitswerte von den Umständen nicht automatisch (wie im 1. Fall) auf die Resultate, sondern wir müssen diesmal die Wahrscheinlichkeitsverteilung für die Resultate R_{ik} , von denen jedes durch h_i und U_k determiniert wird, angeben. Wir treffen hier wieder auf den (uns bereits bekannten) Begriff der bedingten Wahrscheinlichkeit¹⁾, denn wir können diesen Zusammenhang auch so definieren: $P(R_{ik}) = p_{ik} = P(U_k/h_i)$, d.h. $P(R_{ik})$ ist die subjektive Wahrscheinlichkeit dafür, daß das R_{ik} , d.i. der Umstand U_k unter der Voraussetzung, daß die Handlung h_i vollzogen wird, eintritt.

Die Präferenzordnung zwischen Handlungen

Die Aufstellung der drei Matrizen dient dem Zweck, eine Entscheidung als die beste auf eine konsistente und konsequente Weise ableiten zu können, wobei es primär auf die zwei zuletzt angeführten – die Nützlichkeits- und Wahrscheinlichkeitsmatrix – ankommt; die Konsequenzenmatrix dagegen wird gewöhnlich lediglich zwecks der Übersichtlichkeit einer Entscheidungssituation angeführt.

Um sich entscheiden zu können, welche Handlung vollzogen bzw. welche Entscheidung getroffen werden soll, muß das Subjekt die Handlungen bzw. Entscheidungen bewerten. Dies geschieht, indem man den erwarteten Nutzen bzw. den Erwartungswert des Nutzens bestimmt, und zwar auf die folgende Weise:

1) S.15f. dieser Arbeit.

Man multipliziert jede Nützlichkeit in der Zeile einer Nützlichkeitsmatrix, die einer bestimmten Handlung (Entscheidung) h_i entspricht, mit der entsprechenden Wahrscheinlichkeit und addiert anschließend die so erhaltenen Werte miteinander. Sei h_i die Handlung und $u_{i1}, u_{i2}, \dots, u_{in}$ die dieser Handlung entsprechenden Nützlichkeiten aus der Nützlichkeitsmatrix. Ferner seien die entsprechenden Wahrscheinlichkeiten $p_{i1}, p_{i2}, \dots, p_{in}$ gegeben (in einer handlungsunabhängigen Wahrscheinlichkeitsmatrix sind es die gleichen Wahrscheinlichkeiten für alle Handlungen, d.h. $p_1, p_2, \dots, p_n$). Dann ergibt sich der erwartete Nutzen der Handlung h_i (Abk. $E_u(h_i)$):

$$E_u(h_i) = u_{i1}p_{i1} + u_{i2}p_{i2} + \dots + u_{in}p_{in} = \sum_{k=1}^n u_{ik}p_{ik}.$$

Auf diese Weise kann man den erwarteten Nutzen jeder einzelnen der m Handlungen (Entscheidungen) bestimmen und erhält m Erwartungswerte $E_u(h_1), E_u(h_2), \dots, E_u(h_m)$. Die Größenordnung zwischen diesen subjektiven Werten heißt numerische Präferenzordnung der Handlungen (bzw. Entscheidungen).

Die Regel von Bayes

Die Regel von Bayes lautet: Wähle unter den möglichen Handlungen (Entscheidungen) diejenige aus, welche in der numerischen Präferenzordnung den höchsten Wert hat.

Diese Regel empfiehlt also diejenige Handlung als die beste, die die größte Nutzenerwartung für das Subjekt garantiert.

In unserem Beispiel hatten wir folgende Werte: $p = 9/14$, $1-p = 5/14$; $u_{11} = -7$, $u_{12} = -7$, $u_{21} = -2$ und $u_{22} = -16$. Wir erhalten folgende Erwartungswerte des Nutzens:
 $E_u(h_1) = -7 \cdot 9/14 + -7 \cdot 5/14 = -7$;
 $E_u(h_2) = -2 \cdot 9/14 + -16 \cdot 5/14 = -7$.

Das Subjekt kann somit nach der Regel von Bayes jede dieser beiden Handlungen wählen, da jede den größten erwarteten Nutzen für das Subjekt garantiert.

Normierung der Nutzenordnung

Die Nutzenfunktion kann beliebige (positive wie auch negative) reelle Zahlen als Funktionswerte annehmen. Man kann aber die Nutzenwerte stets normieren, d.h. sie auf das Intervall von 0 bis 1 beschränken. Dies ist sogar in der Entscheidungstheorie von Vor-

teil. In diesem Zusammenhang muß das Eindeutigkeitsproblem für die Nutzenordnung gelöst werden. Die Lösung des Eindeutigkeitsproblems besteht im Auffinden einer Funktion ϕ , welche die Nützlichkeitsmatrix in eine äquivalente überführt. Unter äquivalenten Nützlichkeitsmatrizen versteht man dabei solche, die bei gegebenen festen Wahrscheinlichkeiten zur selben qualitativen Präferenzordnung¹⁾ der Handlungen führen.

Es gilt:

- 1) Wenn man jeden Wert der Nützlichkeitsmatrix mit derselben positiven reellen Zahl multipliziert, so entsteht eine äquivalente Nützlichkeitsmatrix.
- 2) Wenn man zu jedem Wert der Nützlichkeitsmatrix eine beliebige reelle Zahl addiert, so entsteht ebenfalls eine äquivalente Nützlichkeitsmatrix.

Die Präferenzordnung zwischen den Handlungen bleibt in beiden Fällen erhalten (leicht nachprüfbar). Das bedeutet: Wenn man jedes u_{ik} der Nützlichkeitsmatrix durch $\alpha u_{ik} + \beta$ mit reellen Zahlen $\alpha > 0$ und β ersetzt, so entsteht eine zu der ursprünglichen Matrix äquivalente Nützlichkeitsmatrix; d.h. die Nützlichkeitsmatrix ist eindeutig höchstens bis auf positiv lineare Transformationen. Die metrische Nutzenordnung (Nutzenskala) ist daher eine Intervallskala²⁾. Dieses Resultat läßt sich in der Weise benutzen, daß wir von nun an die Nützlichkeitsmatrix unabhängig von den Anfangswerten normieren können, d.h. man kann die Nutzenwerte stets so wählen, daß der niedrigste Nutzenwert 0 und der höchste 1 ist.

Das Verfahren sei an einem Beispiel erläutert. Die Nützlichkeitsmatrix (für zwei mögliche Handlungen und zwei mögliche Umstände) sei gegeben als:

1) Sei " $h_i \succ h_j$ " die Abkürzung für " h_i wird h_j nicht vorgezogen". Von einer qualitativen (oder nichtmetrischen) Präferenzordnung der Handlungen spricht man, wenn folgendes gegeben ist: (1) Zwei beliebige Handlungen h_i und h_j sind in bezug auf $\succ$ stets vergleichbar, d.h. es gilt: $h_i \succ h_j$ oder $h_j \succ h_i$; (2) Die Relation $\succ$ ist transitiv, d.h. es gilt für drei beliebige Handlungen h_i, h_j und h_k : Wenn $h_i \succ h_j$ und $h_j \succ h_k$, dann $h_i \succ h_k$. D.h. die qualitative Präferenzordnung der Handlungen ist eine schwache Ordnung.

2) Von einer Intervallskala spricht man, wenn Eindeutigkeit bis auf positive lineare Transformationen vorliegt, d.h. bis auf Transformationen vom Typus $\phi(x) = \alpha x + \beta$ mit positivem reellen α und beliebigem reellen β .

$$\begin{bmatrix} 60 & 6 \\ -30 & 15 \end{bmatrix}$$

Wir ersetzen den niedrigsten Wert -30 durch 0 und den höchsten Wert 60 durch 1. Daraus können wir mittels einer linearen Transformation die anderen Werte berechnen. Wir erhalten die folgenden Gleichungen:

$$(1) \quad 60\alpha + \beta = 1;$$

$$(2) \quad -30\alpha + \beta = 0.$$

Subtrahiert man z.B. (2) von (1), so ergibt sich $\alpha = 1/90$. Durch die Ensetzung dieses Wertes in (1) oder (2) erhält man $\beta = 1/3$. Um die beiden anderen Nützlichkeitswerte zu bestimmen, ersetzt man den rechten oberen Wert 6 durch: $6/90 + 1/3 = 2/5$ und den rechten unteren durch $15/90 + 1/3 = 1/2$. Die transformierte äquivalente Nützlichkeitsmatrix lautet also:

$$\begin{bmatrix} 1 & 0.4 \\ 0 & 0.5 \end{bmatrix}$$

Graphische Lösung des Problems der Maximierung des Erwartungsnutzens

Man kann das Problem der Maximierung des Erwartungswertes des Nutzens auch graphisch lösen, falls nur zwei Umstände (mit $P(U_2) = 1 - P(U_1)$) zur Beurteilung anstehen. Wir machen uns dabei zunutze, daß man den erwarteten Nutzen als Funktion der Wahrscheinlichkeit $P(U_1)$ darstellen kann.

Sei die Nützlichkeitsmatrix wie folgt gegeben:

	U_1	U_2
h_1	1	0.3
h_2	0	0.6

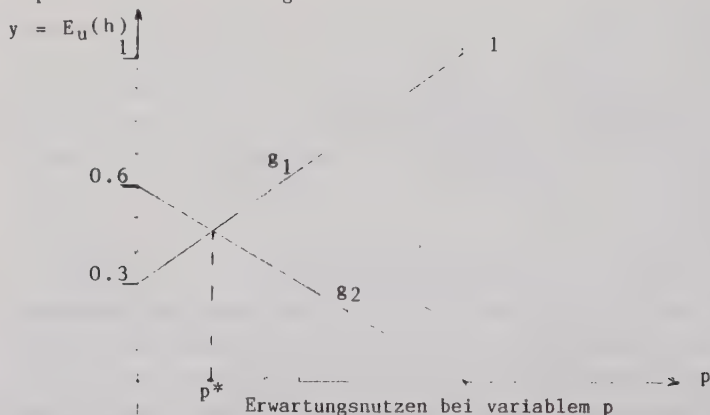
Dann gelten folgende Geradengleichungen:

$$g_1: y = 1 \cdot p + 0.3 \cdot (1-p) = 0.7p + 0.3$$

$$g_2: y = 0 \cdot p + 0.6 \cdot (1-p) = -0.6p + 0.6.$$

(mit $p = P(U_1)$ und $1-p = P(U_2) = 1 - P(U_1)$).

Graphische Darstellung:



Die Abszisse p^* des Schnittpunktes der Geraden g_1 und g_2 ergibt sich aus der Gleichung

$$0.7p^* + 0.3 = -0.6p^* + 0.6$$

$$p^* = 0.2\bar{3}$$

Dies bedeutet, daß die Handlung h_1 optimal wird, wenn die Wahrscheinlichkeit $P(U_1)$ über $0.2\bar{3}$ liegt, dagegen h_2 , wenn $P(U_1)$ unter $0.2\bar{3}$ liegt.

Wir wollen an dieser Stelle die intuitive Einführung in die rationale Entscheidungstheorie vorläufig abschließen mit der Hoffnung, daß der Leser jetzt weiß, um was es sich hier ungefähr handelt. Bevor wir mit der Erläuterung der entscheidungstheoretischen Problematik fortschreiten, werden wir uns im nächsten Paragraphen die bisher von der juristischen Seite vorgeschlagenen Entscheidungsmodelle anschauen. Die in diesem Paragraphen gewonnen Kenntnisse dürfen dabei für das Verständnis dieser Entscheidungsmodelle völlig ausreichen. Im 2. Kapitel wollen wir die bis jetzt eingeführten Begriffe auf die juristische Entscheidungssituation übertragen. Die weitere Entwicklung einer entscheidungstheoretischen Konzeption werden wir dann auf diesen juristischen Rahmen beschränken.

§2. Die bisherigen juristischen Entscheidungsmodelle

Während es an Beiträgen zum Wahrscheinlichkeitsbegriff in der juristischen Literatur nicht mangelt¹⁾, fanden entscheidungstheoretische Argumentationen bis heute wohl kaum Beachtung. Dies ist nicht verwunderlich, denn noch immer versteht sich die Rechtspraxis primär als Gesetzesauslegende und Sachverhaltsklärende, nicht aber als Entscheidungspraxis schlechthin.

Nicht nur von der konservativen Mehrheit wird die Entscheidungsfrage ausgeklammert, sondern auch von denjenigen, die von der Wahrscheinlichkeitstheoretischen Beurteilung juristischer Sachverhalte ausgehen, wird anscheinend nicht klar erkannt, daß diese erst in Hinblick auf die Bildung einer Entscheidung ihre volle Relevanz erfährt, und daher die naheliegende Verbindung von der Wahrscheinlichkeitstheorie zur Entscheidungstheorie, die ein rationales Modell des konsistenten (rationalen) Entscheidens darstellt, völlig übersehen. Die wenigen von der juristischen Seite vorgeschlagenen Entscheidungsmodelle leiden alle an erheblichen Mängeln: entweder sind sie nicht konsistent (Kaplan, Cullison, Maassen) oder nicht adäquat bzw. nicht ausreichend (Marshall&Wise, Motsch, Nagel&Neef).

Was die Entscheidungstheorie selbst betrifft, so wurden hier bis jetzt keine Versuche unternommen, zu überprüfen, ob und inwiefern die in ihrem Rahmen aufgestellten entscheidungstheoretischen Modelle des rationalen menschlichen Verhaltens auch für die Rechtspraxis relevant sein können. Mit dieser Frage werden wir uns übrigens im nächsten Paragraphen auseinandersetzen. Hier geht es uns zunächst darum, die oben angesprochenen juristischen Entscheidungsmodelle zu untersuchen und anhand dieser Analyse (vorläufig informell) herauszustellen, welchen Anforderungen eine adäquate und konsistente juristische Entscheidungstheorie genügen muß.

1) Siehe v.a. I. Teil, 2. Kapitel, aber auch S. 63ff dieser Arbeit.

2.1. Das konservative Entscheidungsmodell

Dieses Modell ist gekennzeichnet durch die folgende Entscheidungsregel¹⁾

Wähle die Verurteilung, wenn der Tatbestand eine ausreichend hohe Wahrscheinlichkeit erreicht bzw. wähle Freispruch, wenn der Tatbestand diese Wahrscheinlichkeit nicht erreicht.

D.h. in diesem Modell determiniert der Sachverhalt vollständig die Entscheidung.

Die obige Entscheidungsregel ist nicht nur (in Form des juristischen Syllogismus) für die herkömmliche Rechtspraxis charakteristisch (wobei hier nicht mehr von einer Wahrscheinlichkeit, sondern vom Bestehen oder Nicht-Bestehen des Sachverhalts die Rede ist), sondern auch für die neuesten Ansätze, die Sachverhaltsfrage wahrrscheinlichkeitstheoretisch anzugehen²⁾

Für die herkömmliche deduktivistische Entscheidungspraxis können wir obige Entscheidungsregel durch folgende differenziertere ersetzen:

Nimm an, daß der Tatbestand, der mit einer ausreichend hohen Wahrscheinlichkeit festgestellt ist, besteht, und wähle - wenn das der Fall ist - die Verurteilung; bzw.: Nimm an, daß der Tatbestand, der nicht mit einer ausreichend hohen Wahrscheinlichkeit festgestellt ist, nicht besteht, und wähle - wenn das der Fall ist - Freispruch.

Der Nachteil dieser Entscheidungsregel besteht darin, daß sie in keiner Weise die Konsequenzen einer Entscheidung berücksichtigt - auch wenn sie für die Betroffenen oder für die Öffentlichkeit noch so verheerend sind.

Außerdem würde die obige Entscheidungsregel - wendet man sie konsequent an - in vielen Fällen zu unsinnigen Entscheidungen führen (wie im folgenden Beispiel):

Angenommen, es sind vier Tatbestände, die beurteilt werden: "Mord", "Totschlag", "minder schwerer Fall von Totschlag" und "Notwehr"; diese seien mit der Wahrscheinlichkeit 0.33, 0.33, 0.33 und 0.01 festgestellt. Keiner der drei Tatbestände "Mord", "Totschlag" und "minder schwerer Fall von Totschlag" besitzt offenbar eine "ausreichend hohe" Wahrscheinlichkeit. Wenn der Richter der obigen Regel folgt, so muß er konsequenterweise "Freispruch" wählen. Angesichts der Tatsache, daß die Wahrscheinlichkeit für "Notwehr" 0.01 beträgt, erscheint diese Entscheidung nicht optimal; denn die Wahrscheinlichkeit dafür, daß der Angeklagte eine Straftat begangen hat (d.h. entweder "Mord" oder "Totschlag" oder "minder schwerer Fall von Totschlag") beträgt immerhin 0.99.

1) Da die dieser Entscheidungsregel zugrunde liegende Verhaltensweise den Juristen anscheinend zu selbstverständlich erscheint, gibt es bis jetzt keinen Versuch, diese Regel zu explizieren bzw. genau zu untersuchen.

2) Vgl. etwa das empirische Modell von Simon & Mahan, S. 150ff dieser Arbeit: in ihrem Modell bestimmt ausschließlich die Wahrscheinlichkeit des Sachverhalts, ob der Angeklagte verurteilt oder freigesprochen wird.

2.2. Das Modell von Kaplan¹⁾

Kaplan kann als erster Rechtswissenschaftler gelten, der versucht hat, ein formales Entscheidungsmodell für die Strafrechtspraxis in Anlehnung an die Entscheidungstheorie zu formulieren²⁾. Leider kann man zeigen, daß Kaplans Modell nicht mit der Entscheidungstheorie konsistent ist und außerdem in vielen Fällen zu unbefriedigenden Ergebnissen führt.

Bei der Einführung seines Modells knüpft Kaplan an den Begriffsapparat der Entscheidungstheorie an. Die zwei Grundbegriffe des Modells sind der subjektive Wahrscheinlichkeitsbegriff³⁾ und der subjektive Nützlichkeitsbegriff.

Kaplan unterscheidet zwischen zwei Entscheidungsalternativen: "Verurteilung" und "Freispruch" und zwischen zwei Umständen: "schuldig" und "nicht schuldig". Außerdem werden zwei Konsequenzen angegeben: "Verurteilung eines Unschuldigen" und "Freispruch eines Schuldigen".

Statt Nützlichkeiten werden den Konsequenzen in Kaplans Modell 'Disutilities' zugeordnet. Die dafür von Kaplan verwendeten Bezeichnungen sind:

D_i : die Disutility, einen Unschuldigen zu verurteilen;

D_g : die Disutility, einen Schuldigen freizusprechen.

Kaplan leitet eine Formel für die Bestimmung der Wahrscheinlichkeit ab, die für den Schuldspruch nötig ist; mit Kaplans Worten (S.1071f):

"Decision theory teaches that in order to convict, the jury must feel that the expected disutility of a decision to acquit is greater than that of a decision to convict. For this to be true, the jury must be convinced that the probability of guilt is at least P , where $P D_g$ is greater than $(1-P) D_i$. Assuming that

1) J.Kaplan, Decision Theory and the Factfinding Process, in: 20 Stanford Law Rev. (1968), S.1065-1092.

2) Vgl. ibid, S.1065: "This Article will attempt to apply some of the basic tools of decision theory to the factfinding process, particularly that of the criminal trial".

3) Allerdings vermengt er - wie es oft in der juristischen Literatur der Fall ist - den subjektiven Wahrscheinlichkeitsbegriff mit dem Begriff der relativen Häufigkeit, wenn er auf S.1073 schreibt: "First and most obvious, if we tell the jury that they must be 99.5-percent certain of the guilt of the accused before they convict him, we are telling them that they will and indeed are expected to convict 1 innocent man in each 200 guilty verdicts".

$D_i + D_g > 0$, P can then be expressed in terms of the different disutilities:

$$P > \frac{1}{1 + \frac{D_g}{D_i}} \quad " 1)$$

Wir wollen nun sehen, wieso das Modell von Kaplan nicht funktioniert. Wir knüpfen an die auf S.242f dieser Arbeit abgeleitete Konsequenzenmatrix an.

Da im obigen Fall zwei Umstände und zwei Entscheidungsalternativen vorliegen, sind insgesamt vier Resultate möglich. Die entsprechenden Konsequenzenmatrix ist somit:

	U_1	U_2
h_1	R_{11}	R_{12}
h_2	R_{21}	R_{22}

Im obigen Fall lautet die Entscheidungssituation konkret:

	schuldig	unschuldig
Verurteilung	Verurteilung eines Schuldigen	Verurteilung eines Unschuldigen
Freispruch	Freispruch eines Schuldigen	Freispruch eines Unschuldigen

Die entsprechende Nützlichkeitsmatrix²⁾ muß dann (mit dis_{ij} , $i, j = 1, 2$ für die entsprechenden Disutilities) folgendermaßen lauten:

$$\begin{bmatrix} dis_{11} & dis_{12} \\ dis_{21} & dis_{22} \end{bmatrix}$$

In Kaplans Modell sieht sie dagegen so aus (mit $dis_{12} \hat{=} D_i$ und $dis_{21} \hat{=} D_g$):

$$\begin{bmatrix} - & dis_{12} \\ dis_{21} & - \end{bmatrix}$$

1) Anzumerken ist folgendes: die letzte Ungleichung gilt, wenn sowohl D_i wie auch D_g größer als 0 ist. Wenn D_i und D_g kleiner als 0 sind, dann ist statt ">" das Zeichen "<" zu verwenden.

2) Genauer müßte diese Matrix hier "Disutility"- bzw. "Verlust"-Matrix heißen. Aus Gründen der Einfachheit und Übersichtlichkeit wollen wir jedoch den alten Terminus "Nützlichkeitsmatrix" beibehalten, zumal der Begriff der Nützlichkeit (und nicht der Begriff "disutility") der Grundbegriff der rationalen Entscheidungstheorie ist.

einzig und allein von den entsprechenden Nützlichkeiten (bzw. Disutilities) ab, da die Wahrscheinlichkeit für den zweiten Tatbestand völlig durch die des ersten determiniert ist¹⁾. Dieses Modell erfaßt nicht Fälle mit mehr als 2 Umständen und (entsprechend) mit mehr als 2 Entscheidungsmöglichkeiten. In dem komplexeren Modell ist die Ableitung der zur Verurteilung ausreichenden Wahrscheinlichkeit aus den Nützlichkeiten nicht mehr möglich, denn sie hängt nicht nur von diesen, sondern auch von den Wahrscheinlichkeiten für die anderen Umstände ab. Diese Wahrscheinlichkeiten sind dann natürlich nicht mehr durch die Wahrscheinlichkeit für den in Frage kommenden Tatbestand determiniert²⁾.

2. Im Kaplan-Modell ist auch dann die Verurteilung ableitbar, wenn die Wahrscheinlichkeit für "Schuld" kleiner ist als eine solche für "Unschuld". Dies widerspricht eindeutig der juristischen Gepflogenheit, nur dann zu verurteilen, wenn die Wahrscheinlichkeit für "Schuld" hoch bzw. sehr hoch (auf jeden Fall aber größer als diejenige für "Unschuld") ist.

3. Das Modell gibt keine Antwort auf die Frage, was zu tun ist, wenn beide Entscheidungsmöglichkeiten den gleichen Erwartungswert erhalten. Zwar kann man in der Entscheidungstheorie in einem solchen

1) Falls man Disutilities nicht nur zwei, sondern allen möglichen Resultaten R_{ij} , $i, j=1,2$, zuordnet, sieht die Formel für die Ableitung der für die Verurteilung notwendigen Wahrscheinlichkeit so aus:

$$P > \frac{1}{1 + \frac{\text{dis}(R_{21}) - \text{dis}(R_{11})}{\text{dis}(R_{12}) - \text{dis}(R_{22})}}$$

2) Kaplan untersucht auch einen solchen Fall (vgl. S. 1079f), versucht ihn aber auf den Fall der 2×2 -Matrix zu reduzieren, indem er die Wahrscheinlichkeiten für die übrigen $n-2$ Umstände von vornherein festlegt. Offensichtlich meint auch Kaplan, daß keine allgemeine Regel für die Ableitung der Wahrscheinlichkeit für den in Frage kommenden Tatbestand aus den Nützlichkeiten im Fall einer $n \times n$ -Matrix angebar ist, denn auf S. 1079 kehrt er zu der klassischen Entscheidungsregel der Maximierung des Erwartungsnutzens (bzw. bei ihm der Disutilityminimierung) zurück.

Fall jede Entscheidung wählen. Dies ist jedoch in vielen juristischen Fällen sehr unbefriedigend, etwa dann, wenn z.B. die Verurteilung wegen Mordes und Freispruch im Fall der Notwehr den gleichen maximalen Erwartungswert des Nutzens besitzen.

Das Modell von Kaplan ist ferner in folgender Hinsicht unbefriedigend:

- 1) Kaplan macht sich überhaupt keine Gedanken über die Nutzen- bzw. Disutility-Funktion. Wir erfahren nicht, was Kaplan sich unter "Nutzen" ("disutility") vorstellt, insbesondere, ob es sich dabei um einen sozialen oder persönlichen Nutzen handelt¹⁾
- 2) Falls die Anzahl der Umstände größer als 2 ist, so kann man bei Kaplan auch dann eine Entscheidung entsprechend einem bestimmten Tatbestand als die beste ableiten, wenn dieser Tatbestand die Wahrscheinlichkeit 0 hat. Dies ist aber höchst unbefriedigend, wenn es sich bei diesem Tatbestand um einen strafbaren handelt²⁾

Abschließend soll hier noch die Kritik von Tribe³⁾ an Kaplan kurz erwähnt werden. Tribe nimmt u.a. an, daß die Nützlichkeiten der Konsequenzen in juristischen Fällen von den Wahrscheinlichkeiten für die Umstände abhängen müßten und daher eine juristische Entscheidungssituation durch ein Kaplan-Modell - bzw. allgemeiner, durch ein entscheidungstheoretisches Modell - nicht adäquat dargestellt werden könnte.

Würden die Konsequenzen "Der Angeklagte wird verurteilt" bzw. "Der Angeklagte wird freigesprochen" lauten, so wäre Tribe's Kritik berechtigt. Denn es ist plausibel, daß je höher (kleiner) die Wahrscheinlichkeit für die Täterschaft dieses Angeklagten ist, desto wünschenswerter (weniger wünschenswert) ist, daß dieser Angeklagte verurteilt (freigesprochen) wird. Allerdings sind das nicht die Konsequenzen, die in juristischen Fällen entscheidungstheoretisch bewertet werden. Diese lauten (ausführlich formuliert): "Der Angeklagte ist der Täter und der Angeklagte wird verurteilt" (oder kurz: "Der Täter wird verurteilt"), "Der Angeklagte ist der Täter und der Angeklagte wird nicht verurteilt" (kurz: "Der Täter wird nicht verurteilt"), "Der Angeklagte ist kein Täter und der Angeklagte wird verurteilt" (kurz: "Ein Nicht-Täter wird verurteilt") und "Der Angeklagte ist kein Täter und der Angeklagte wird nicht verurteilt" (kurz: "Ein Nicht-Täter wird nicht verurteilt").

Es ist schwer einzusehen, wieso die Wünschbarkeit dieser Konsequenzen von der Wahrscheinlichkeit für die Täterschaft des Angeklagten abhängen soll (oder konkreter: wieso es z.B. weniger wünschenswert ist, den Täter zu verurteilen, wenn man glaubt, daß der Angeklagte die Tat gar nicht begangen hat)⁴⁾

1) Vgl. auch die Kritik von P.G.Milanich, *Decision Theory and Standards of Proof*, in: 5 *Law and Behavior* (1981), S.87-96.

2) Vgl. Kaplan, S.1080.

3) L.Tribe, *Trial by Mathematics*, S.1378-1393.

4) Der erwartete Nutzen der Handlungen "Der Angeklagte wird verurteilt" bzw. "Der Angeklagte wird freigesprochen" hängt hingegen von der Wahrscheinlichkeit für die Umstände ab. Dies ist aber trivial. Zu einer weiteren Kritik an Tribe siehe Milanich (Fn.1 oben).

Obwohl das Modell von Kaplan - wie wir gesehen haben - inkonsistente und inadäquate Ergebnisse liefert, übernahm es im deutschen Sprachraum Maassen kritiklos, um seine Theorie des Beweismaßes darauf aufzubauen.

2.3. Das Modell von Maassen¹⁾

Maassen formuliert unter dem Gesichtspunkt der Kostenminimierung, die aus der fehlerhaften Feststellung bzw. der fehlerhaften Nichtfeststellung von Tatsachen resultiert, ein Entscheidungsmodell, das demjenigen Kaplans ähnlich ist. Die Kosten werden dabei von Maassen im Sinn des individuellen und sozialen Schadens verstanden²⁾

Leider ist die Ausdrucksweise Maassens so unpräzise (er verwendet die gleichen Ausdrücke "fehlerhafte Feststellung/ fehlerhafte Nichtfeststellung von Tatsachen" sowohl in bezug auf Konsequenzen als auch in bezug auf die Umstände; seine Handlungen lauten wiederum "Feststellung/Nichtfeststellung"), so daß es schwierig ist, die von Maassen anvisierte Entscheidungssituation zu rekonstruieren, d.h. man weiß u.a. nicht, welchen Entitäten im Modell Maassens welcher Status - Handlung, Umstand oder Konsequenz - zukommt. Die unten angegebene Entscheidungssituation gibt vielleicht Maassens Intention besser wieder (wobei ich die Ausdrucksweise Maassens nur dort beibehalten habe, wo diese zu keinen großen Mißdeutungen führen kann):

Handlungen/Umstände	Tatsachen liegen vor	Tatsachen liegen nicht vor
Feststellung	-	Tatsachen werden festgestellt, obwohl sie nicht vorliegen ("fehlerhafte Feststellung von Tatsachen")
Nichtfeststellung	Tatsachen werden nicht festgestellt, obwohl sie vorliegen ("fehlerhafte Nichtfeststellung von Tatsachen")	-

(Maassens Abkürzungen für "fehlerhafte Feststellung von Tatsachen" und "fehlerhafte Nichtfeststellung von Tatsachen" sind jeweils "ff" und "fFN").

Wie wir sehen, interessiert sich Maassen ähnlich wie Kaplan nur für die beiden ungünstigen Konsequenzen.

¹⁾ B.M.Maassen, Beweisprobleme im Schadenersatzprozeß (1975), S.5ff.
²⁾ Vgl. ibid, S.7.

Benutzen wir die Abkürzungen von Maassen ($K_{fF} \hat{=}$ "Kosten einer fehlerhaften Feststellung von Tatsachen", $K_{fNF} \hat{=}$ "Kosten einer fehlerhaften Nichtfeststellung von Tatsachen"), so erhalten wir die der obigen Konsequenzenmatrix entsprechende Nützlichkeitsmatrix (bzw. "Kostenmatrix"):

$$\begin{bmatrix} - & K_{fF} \\ K_{fNF} & - \end{bmatrix}$$

Da das Ziel in der Kostenminimierung liegt, soll man diejenige Handlung wählen, deren Erwartungswert bzgl. der Kosten am geringsten ist; d.h. man soll die "Feststellung" nur dann wählen, wenn die erwarteten Kosten der "fehlerhaften Feststellung von Tatsachen" geringer sind als die erwarteten Kosten der "fehlerhaften Nichtfeststellung von Tatsachen".

Ist P die Wahrscheinlichkeit dafür, daß die Tatsachen vorliegen, so lautet die Entscheidungsregel von Maassen für die Wahl der ersten Handlung ("Feststellung"):

$$K_{fNF}P(fNF) > K_{fF}P(fF) = K_{fF}(1 - P(fNF)).$$

Daraus ist die für die Wahl dieser Handlung notwendige Wahrscheinlichkeit für das Vorliegen der Tatsachen ableitbar:

$$P = P(fNF) > \frac{1}{1 + \frac{K_{fNF}}{K_{fF}}}$$

Setzen wir in diese Formel bzw. in die obige Nützlichkeitsmatrix anstelle von K_{fF} und K_{fNF} die Disutilities dis_{12} und dis_{21} ein, so erhalten wir die uns bereits aus dem Modell von Kaplan bekannte Entscheidungsregel:

$$P > \frac{1}{1 + \frac{dis_{21}}{dis_{12}}}$$

Somit gelten alle dort vorgebrachten Einwände.

Positiv ist allerdings bei Maassen anzumerken, daß er im Fall der gleichen Erwartungswerte der Kosten für die beiden Alternativen an die juristische Praxis anknüpft, zuungunsten desjenigen zu entscheiden, der den Status Quo zu seinem Vorteil verändern will, d.h. des Klägers (vgl. S.10).

2.4. Das Modell von Motsch¹⁾

Motsch formuliert sein Modell unter dem Gesichtspunkt der Minimierung der Schädlichkeit, die sich aus den Folgen einer etwaigen Fehlentscheidung ergibt.

Motsch unterscheidet zwischen 2 Umständen "schuldig"/"nicht schuldig" und 2 Entscheidungsalternativen "Verurteilung"/"Freispruch". Im Unterschied zu Kaplan und Maassen gibt er auch die vier möglichen Konsequenzen richtig wieder. Diese sind:

1. A ist schuldig und wird verurteilt;
2. A ist schuldig und wird nicht verurteilt;
3. A ist nicht schuldig und wird verurteilt;
4. A ist nicht schuldig und wird nicht verurteilt.

Allerdings setzt Motsch die Schädlichkeit der Konsequenzen 1 und 4 gleich 0 (somit gilt die von mir auf S.255 vorgetragene Kritik).

Mit der Abkürzung P für die Wahrscheinlichkeit, daß A schuldig ist, dis_{12} für die Konsequenz 3 und dis_{21} für die Konsequenz 2, ferner mit

$$e = \frac{1}{1 + \frac{dis_{21}}{dis_{12}}}$$

lautet die Entscheidungsregel von Motsch (S.335):

wenn $P < e$, dann Freispruch

wenn $P \geq e$, dann Schuldspruch.

Wie wir sehen gibt Motsch im Falle des Schuldspruchs die gleiche Entscheidungsregel wie Kaplan bzw. Maassen an, mit der Ausnahme²⁾, daß er für den Fall, daß beide Entscheidungsalternativen den gleichen Erwartungswert aufweisen, die Wahl der härteren Entscheidung "Schuldspruch", d.h. die Verurteilung, empfiehlt. (Wie wir bereits wissen, hat Kaplan für diesen Fall überhaupt keine Regel formuliert, Maassen dagegen die Wahl der milderer Entscheidung empfohlen).

Zur weiteren Kritik an Motsch siehe die nachfolgende Diskussion des Modells von Marshall&Wise, das dem von Motsch im wesentlichen gleicht, so daß die dort geäußerte Kritik auch hier gilt.

1) R.Motsch, Vom Prozeß als Beweis zum Überwiegensprinzip, in: Gedächtnisschrift für J.Rödig (1978), S.334-348.

2) Eine weitere Ausnahme ist, wie oben erwähnt, die, daß er auch die Konsequenzen 1 und 4 in Betracht zieht. Da er ihnen allerdings den Wert 0 zuordnet und außerdem die Zuordnung von Werten zu beiden anderen Konsequenzen auf die positiven reellen Zahlen einschränkt, erhalten wir den gleichen Ausdruck auf der rechten Seite der Ungleichung wie bei Kaplan (d.h. $1/(1+dis_{21}/dis_{12})$).

2.5. Das Modell von Marshall&Wise

Das Modell von Marshall&Wise wurde bereits auf S.157f dieser Arbeit vorgestellt (im Zusammenhang mit der Schilderung ihrer empirischen Untersuchung der Frage, inwiefern Individuen sich in der Praxis entsprechend dem Bayesianistischen Entscheidungsmodell verhalten). Wir wollen dieses Modell hier noch einmal kurz darstellen.

Das Entscheidungsmodell von Marshall&Wise geht von zwei Umständen "Schuld"/"Unschuld" und zwei Entscheidungsalternativen "schuldig"/"unschuldig" aus. Es unterscheidet zwischen vier Konsequenzen; so gesehen ist dieses Modell also konsistent. Das abstrakte Schema ihrer Nützlichkeitsmatrix sieht so aus:

$$\begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix}$$

(Anstelle der Bezeichnungen von Marshall&Wise benutzen wir hier unsere Abkürzungen; so korrespondiert u_{11} mit U_{hit} , u_{12} mit $U_{f.a.}$, u_{21} mit U_{mis} , u_{22} mit $U_{c.r.}$; vgl. S.158 dieser Arbeit).

Allerdings ordnen Marshall&Wise den drei Konsequenzen feste Nützlichkeiten zu, d.h. $u_{11} = u_{22} = 0$, $u_{21} = -1$, und lassen nur u_{12} variieren, d.h. $u_{12} = -x$.

Ist P die Abkürzung für die Wahrscheinlichkeit der Schuld des Angeklagten, so lautet Marshall&Wise's Entscheidungsregel wie folgt:

Um die Entscheidung "schuldig" abzuleiten, muß gelten

$$u_{11}P + u_{12}(1-P) \geq u_{21}P + u_{22}(1-P)^1)$$

Hält man - wie oben angegeben - u_{11} , u_{22} und u_{21} konstant, so gilt für P :

$$P \geq \frac{1}{1 - \frac{1}{u_{12}}} = \frac{1}{1 + \frac{1}{x}}$$

Trotz seiner Konsistenz ist das Modell von Marshall&Wise (ebenso wie das von Motsch) in dem Sinn inadäquat, als auch hier die gegen das Modell von Kaplan vorgebrachten Kritikpunkte 1 und 2 (vgl. S.255f dieser Arbeit) in vollem Umfang weitergelten. Außerdem ist es wie

1) Die Entscheidungsregel, die Marshall&Wise vorschlagen, besteht also in der Nutzenmaximierung.

alle vorhergehenden Modelle unbefriedigend, da es den verwendeten Grundbegriff "Nützlichkeit"¹⁾ inhaltlich nicht klärt.

Weiter ist das Modell von Marshall&Wise zu kritisieren, weil es für den Fall, daß beide Entscheidungen den gleichen Erwartungswert haben, (ebenso wie das Modell von Motsch) die Wahl der härteren Alternative ("schuldig") vorschlägt, was der (humanen) juristischen Forderung "im Zweifel für den Angeklagten" widerspricht. Nicht zuletzt kann man sich auch fragen, inwiefern ihre willkürliche Apriori-Festlegung der Nützlichkeiten befriedigend ist, denn dies bringt die beurteilende Person von vornherein um die Möglichkeit, frei nach ihren Vorstellungen bzgl. der Konsequenzen zu entscheiden. (Bewertet das Individuum jedoch die Konsequenzen nicht im Einklang mit den von Marshall&Wise festgelgten Nützlichkeiten, so gilt die oben (auf Grund dieser Festlegungen) abgeleitete und für ihre empirische Untersuchung maßgebende²⁾ Formel für P nicht mehr bzw. muß diese Formel durch eine kompliziertere ersetzt werden).

1) Bei Motsch sind es "Schädlichkeiten", vgl. S.260 dieser Arbeit (zur Erinnerung: bei Kaplan sind es "disutilities" und bei Maassen "Kosten" im Sinn von "Schaden").

2) Vgl. S.158 dieser Arbeit.

2.6. Das Modell von Cullison¹⁾

Cullisons Modell ist im Wesentlichen dasjenige von Kaplan: So unterscheidet er zwischen 2 Umständen: "Schuld"/"Unschuld", 2 Entscheidungsalternativen: "Verurteilung"/"Freispruch" und nur zwei unerwünschten Konsequenzen: "Verurteilung eines Unschuldigen" und "Freispruch eines Schuldigen".

Cullison versucht - wie bereits Maassen, Motsch und Marshall&Wise - nicht nur eine Entscheidungsregel zu formulieren, sondern sie bereits im voraus sozusagen apriori numerisch festzulegen. So schreibt er (S.238):

"From the standpoint of the Model of Rules, a jury can make two kinds of mistake: (I) it can convict a defendant who is actually innocent (not guilty) under the rules, or (II) it can acquit a defendant who is actually guilty. Probability theorists would call these Type I und II errors, respectively. There is a definite connection between the threshold probability (degree of belief) which should rationally trigger a conviction and the relative desirabilities (presumably negative quantities) of these two types of error. The connection is implicit in a statement like this: "It is better to let 1,000 guilty men go free than to convict a single innocent man". In other words, it is more desirable to make 1,000 Type II errors than it is to make a single I Type error, or $D(I) < 1,000 \times D(II)$, where $D(I)$ represents the desirability of a Type I error and $D(II)$ the desirability of a Type II error".

Da die Nützlichkeiten (bzw. Wünschbarkeiten) der Konsequenzen sowohl durch positive wie auch durch negative reelle Zahlen repräsentiert werden können (die Nutzentheorie kennt diesbezüglich keine Einschränkungen²⁾), können wir den obigen Satz "It is better to let 1.000 guilty man go free than to convict a single innocent man" durch folgende Nützlichkeitsmatrix z.B. wiedergeben:

$$\begin{bmatrix} - & 1 \\ 1000 & - \end{bmatrix}$$

Mit dieser Matrix erhalten wir das unsinnige Ergebnis, daß "Verurteilung" für sehr geringe Wahrscheinlichkeitswerte für P ($< 1/1001$) und "Freispruch" für hohe Werte für P (u.a. sogar für $P = 1$) gewählt werden soll.

Es ist klar, daß eine solche Nützlichkeitsmatrix kaum brauchbar ist, denn die Konsequenzen der - sozial gesehen - wünschenswerten Konsequenzen: "Verurteilung eines Schuldigen"/"Freispruch eines Unschul-

1) A.D.Cullison, The Model of Rules and The Logic of Decision, in: Nagel, S.S. (Hrsg.), Modeling the Criminal Justice System (1977), S. 225-246.

2) Cullison schreibt zwar im obigen Zitat, daß die Nützlichkeiten "vermutlich" durch negative Zahlen zu repräsentieren seien ("the relative desirabilities (presumably negative quantities)"); da die Nutzentheorie diese Einschränkung nicht kennt, wäre es aber wünschenswert gewesen, wenn Cullison - was er leider nicht getan hat - diese Einschränkung näher erklärt bzw. begründet hätte.

digen" kommen darin nicht vor, im Gegensatz zu den unerwünschten Konsequenzen, denen trotzdem Nützlichkeiten (desirabilities) im Cullison-Modell zugeordnet werden müssen. Die Werte 1000 für "Freispruch eines Schuldigen" und 1 für "Verurteilung eines Unschuldigen" erhalten offensichtlich erst dann einen Sinn, wenn wir auch den restlichen Konsequenzen numerische Werte zuordnen, etwa wie in der nachfolgenden Nützlichkeitsmatrix:

$$\begin{bmatrix} 10^4 & 1 \\ 10^3 & 10^5 \end{bmatrix}$$

Diese Nützlichkeitsmatrix können wir positiv linear transformieren¹⁾, so daß wir eine dazu äquivalente Nützlichkeitsmatrix erhalten, die so aussieht:

$$\begin{bmatrix} 1/9 & 0 \\ 1/99 & 1 \end{bmatrix}$$

Mit dieser Nützlichkeitsmatrix wird die Verurteilung gewählt, wenn die Wahrscheinlichkeit für "Schuld" ≥ 0.9 ist.

1) Im Modell von Kaplan (Motsch, Cullison) sind dagegen positive lineare Transformationen, d.h. Transformationen vom Typus $\phi(x) = \alpha x + \beta$ mit positivem reellen α und beliebigem reellen β , nicht möglich. Dagegen gelten Ähnlichkeitstransformationen, d.h. Transformationen von der Gestalt $\phi(x) = \alpha x$ mit positivem reellen α .

2.7. Das Modell von Nagel&Neef¹⁾

Auch hier²⁾ handelt es sich um ein einfaches und konsistentes Entscheidungsmodell, das nur 2 Umstände: "Schuld"/"Unschuld", 2 Entscheidungsalternativen: "Verurteilung"/"Freispruch" und die 4 uns bekannten Konsequenzen kennt. Den Konsequenzen werden subjektive Nützlichkeiten zugeordnet, die Nagel&Neef als "moralische Genugtuung" interpretieren. Die Zuordnung von Nützlichkeiten zu den Konsequenzen wird - im Unterschied zu den bisherigen Modellen - keinen Einschränkungen unterworfen, so daß die Entscheidungsregel, die in der Maximierung des Erwartungsnutzens besteht, hier ihre volle Anwendung findet. Allerdings kann in dem Fall, daß die beiden Entscheidungsalternativen den gleichen Erwartungswert haben, jede Entscheidung als die beste gewählt werden. Die willkürliche Wahl kann bei Nagel&Neef etwa dadurch gelöst werden, daß man eine normale Münze wirft. (Wie bereits erwähnt, ist dies jedoch höchst unerwünscht, wenn es sich um in ihren Konsequenzen stark voneinander divergierenden Entscheidungen handelt (wie z.B. diejenigen, die auf Mord" einerseits und "Notwehr" andererseits folgen)). Darüber hinaus ist das eigentliche Ziel auch dieses Entscheidungsmodells nicht die Ableitung einer Entscheidung (anhand den Wahrscheinlichkeiten für die Sachverhalte und den Nützlichkeiten für die jeweiligen Konsequenzen), sondern die eines Wahrscheinlichkeitswertes (aus den Nützlichkeiten und der Präferenzordnung der Entscheidungen), die für die Verurteilung ausreicht. Das funktioniert (wie bereits in der Kritik am Modell Kaplans dargestellt wurde) nur in einem solchen, lediglich 2 Umstände berücksichtigenden Modell. Ferner trägt eine derartige Vorgehensweise auch dazu bei, das Ergebnis zu verzerren (wie auch von Nagel&Neef beklagt wird³⁾), denn man geht hier bereits von einer vorgegebenen Präferenzordnung aus, anstatt diese

1) S.S.Nagel&M.G.Neef, Decision Theory and the Legal Process (1979).

2) Vgl.Nagel&Neef, S.188f.

3) Nagel&Neef, S.199f: "Jurors often may decide first whether to convict or acquit and then an appropriate probability of guilt and a conviction standard".

erst im nachhinein anhand der festgestellten Wahrscheinlichkeiten und den unabhängig davon bestimmten Nützlichkeiten abzuleiten¹⁾).

Auch für das Modell von Nagel&Neef gilt im vollen Umfang die Kritik, die bereits in Verbindung mit anderen Modellen wegen deren Nichtadäquatheit vorgebracht wurde.

1) Nagel&Neef stellen sich u.a. (angesichts des in Fn.3 auf der vorangehenden Seite angedeuteten Problems) die Frage: "Those findings raise the question of...how jurors can be better instructed so as to increase the likelihood that they will consciously think about the defendant's probability of guilt and the conviction standard before, not after, deciding how to vote". Eine Möglichkeit, das wenigstens zum Teil zu garantieren, besteht vermutlich darin, daß der Richter öffentlich zunächst seine Nützlichkeitsmatrix und unabhängig davon die Wahrscheinlichkeiten für die in Frage kommenden Sachverhalte angibt, so daß erst im nachhinein die Entscheidung abgeleitet wird. Da der Entscheidungsprozeß öffentlich kontrolliert wird, kann der Richter dann nicht - wenn ihm die rational abgeleitete Entscheidung mißfällt - schnell seine Nützlichkeiten bzw. Wahrscheinlichkeiten in der Weise verändern, um die von ihm gewünschte Entscheidung doch noch abzuleiten. Es ist klar, daß auch dann das Problem nicht völlig gelöst ist, denn auch unter solchen Bedingungen bleibt genug Raum für unehrliche Manipulationen. Allerdings ist die Vorgehensweise, zuerst Wahrscheinlichkeiten und Nützlichkeiten festzustellen und dann die Entscheidung abzuleiten, auf jeden Fall derjenigen vorzuziehen, die darin besteht, aus der Entscheidungspräferenz die notwendige Wahrscheinlichkeit für den Sachverhalt abzuleiten, denn im ersten Fall entscheidet nicht die hypothetische bzw. mitunter sogar die tatsächliche Entscheidungspräferenz (etwa, wenn der Richter den Angeklagten unbedingt freisprechen will) über den Sachverhaltsaufklärungs- und Entscheidungsprozeß und dessen Ausgang, sondern einzig und allein die festgestellten Wahrscheinlichkeiten für die Sachverhalte und die Nützlichkeiten für die jeweiligen Konsequenzen.

Zusammenfassung und Ausblick

In diesem Paragraphen haben wir gezeigt, daß alle bisher von der juristischen Seite vorgeschlagenen Entscheidungsmodelle inadäquat bzw. nicht ausreichend sind, und daß einige davon (exemplarisch dafür kann das Modell von Kaplan gelten) noch darüber hinaus mit der Entscheidungstheorie nicht konsistent sind. Alle Modelle leiden darunter, daß sie die denkbar einfachste Entscheidungssituation darstellen, wobei das mit der Aufstellung solcher Modelle verfolgte Ziel eigentlich nicht in der Ableitung einer rationalen Entscheidung, sondern in der einer Wahrscheinlichkeit für den Tatbestand, die für eine bestimmte Entscheidung ausreicht, gesehen wird; darüber hinaus werden in der Regel Nützlichkeiten willkürlich eingeschränkt, ohne das durch irgendwelche Regeln zu sichern.

Demgegenüber wollen wir im folgenden ein sowohl mit der Entscheidungstheorie konsistentes wie auch gegenüber der Rechtspraxis adäquates juristisches Entscheidungsmodell konstruieren. Aus der an den einzelnen juristischen Entscheidungsmodellen geübten Kritik ergeben sich folgende Bedingungen, denen dieses Modell genügen muß:

- 1) Es sollen darin beliebig (endlich) viele Tatbestände und dementsprechend beliebig (endlich) viele Entscheidungsmöglichkeiten repräsentiert werden.
- 2) Es soll eine juristisch konforme Lösung enthalten für den Fall, daß mehrere Entscheidungsalternativen den höchsten Erwartungswert des Nutzens erhalten.
- 3) In diesem Modell soll wenigstens garantiert werden, daß die Wahrscheinlichkeit des Sachverhalts, der der besten abgeleiteten Entscheidung zugrunde liegt, nicht am geringsten ist bzw. nicht unter eine gewisse untere Schranke fällt¹⁾ (Dadurch werden u.a. auch bestimmte unerwünschte Ergebnisse ausgeschaltet, die man bis jetzt dadurch zu vermeiden suchte, daß man den Konsequenzen nur bestimmte - aber nichtsdestoweniger willkürliche - Nützlichkeiten zuordnete).

Außerdem wollen wir uns über folgende Gesichtspunkte Klarheit verschaffen:

- 1) Der Begriff der subjektiven Nützlichkeit, neben dem Wahrscheinlichkeitsbegriff, der zweite wichtigste Grundbegriff der Entscheidungstheorie, speziell der des juristischen Nutzens, soll erörtert werden. Dabei wollen wir berücksichtigen, daß es sich bei diesem nicht um einen persönlichen Nutzen (auch wenn er subjektiv bestimmt wird) handelt, sondern um einen sozialen, der (unter dem Oberbegriff "juristische Zielsetzungen und -vorstellungen") solche Aspekte umfaßt wie Gerechtigkeit, Wiedereingliederung in die Gesellschaft, Wiedergutmachung des verschuldeten Unrechts, Verbrechensverhütung und -bekämpfung, aber auch (im Rahmen des juristisch Erlaubten) Vergeltung, Genugtuung der Geschädigten, Sicherheitsbedürfnis der Gesellschaft (mittels Präventionen), angemessene Leidensverringerung des Bestraften.
- 2) Zwischen der Täterschaft (objektiver Tatbestand) und der Schuld (subjektiver Tatbestand) muß offensichtlich scharf differenziert werden. In den bisherigen Modellen wurde eine solche Unterscheidung

¹⁾ Wie gezeigt wurde, erfüllt keines der bisherigen Entscheidungsmodelle diese Bedingung.

nicht gemacht; es wurde also nicht berücksichtigt, daß die Wahrscheinlichkeit für die Schuld (die neben dem empirischen auch den psychologischen Aspekt umfaßt) nicht gleich der Wahrscheinlichkeit für den Sachverhalt gesetzt werden kann. Die deutsche Rechtsprechung knüpft die gesetzlichen Rechtsfolgen an die objektiven Tatbestände an. Erst der Vollzug bzw. die Bestätigung der Entscheidung hängt davon ab, ob der Angeklagte schuldig bzw. schuldfähig ist oder nicht. Insbesondere wird das klar, wenn wir mit einer Entscheidungssituation konfrontiert werden, wo mehrere Tatbestände (z.B. "Mord", "Totschlag", "fahrlässige Tötung") und entsprechend mehrere unterschiedliche Entscheidungsmöglichkeiten zur Auswahl stehen. Wird aus diesen drei Tatbeständen etwa derjenige des Mordes mit der höchsten Wahrscheinlichkeit festgestellt und darüber hinaus die diesem Tatbestand entsprechende Entscheidung "lebenslänglich" abgeleitet, so hängt offensichtlich der Vollzug bzw. die endgültige Bestätigung der Entscheidung davon ab, ob der Angeklagte schuldfähig ist oder nicht.

3) Es muß geklärt werden, um welche Entscheidungssituation es sich bei juristischen Entscheidungen überhaupt handelt. Wie wir gesehen haben, hat man bis jetzt in allen Modellen (mit Ausnahme des konventionellen Entscheidungsmodells) es als selbstverständlich angenommen, daß hier eine bestimmte Entscheidungssituation vorliegt. Diese enthält drei Arten von Entitäten: Umstände, Entscheidungen und Konsequenzen, die durch Umstände und Entscheidungen determiniert sind; Umständen werden Wahrscheinlichkeiten, Konsequenzen werden Nützlichkeiten (oder Disutilities) zugeordnet und für die möglichen Entscheidungen Erwartungswerte abgeleitet. Kurz, man nimmt (ohne es zu wissen) an, daß es sich bei den juristischen Entscheidungsmodellen um sog. "Savage-Modelle" handelt. Aus der Entscheidungstheorie sind neben Savage-Modellen aber mindestens noch zwei andere Arten von Entscheidungsmodellen bekannt, nämlich die Jeffrey- und die Fishburn- Modelle, die gewisse Entscheidungssituationen adäquater wiedergeben als diejenigen von Savage. Es muß somit gezeigt werden, daß juristische Entscheidungs-Modelle in der Tat Savage-Modelle sind (Mit dieser Frage werden wir uns eingehend im nächsten Paragraphen befassen).

4) Und nicht zuletzt muß Klarheit über den Begriffsapparat einer juristischen Entscheidungstheorie gewonnen werden. Die juristische Uminterpretation entscheidungstheoretischer Begriffe ist nicht nur deshalb notwendig, damit der Jurist sich in einem ihm (terminologisch bekannten) Bereich vorfindet, sondern auch, um die Einheitlichkeit der Begriffsbildung (die bis jetzt in den verschiedenen Modellen zu kurz gekommen ist) zu garantieren. (Mit der begrifflichen Klärung einer juristischen Entscheidungssituation beginnen wir das nächste Kapitel).

Abschließend soll hier noch ein Gesichtspunkt erwähnt werden, der bisher von keinem der untersuchten juristischen Entscheidungsmodelle (und auch noch nicht in dieser Arbeit) behandelt wurde, nämlich der, daß die Entscheidungsbildung nicht nur in der Wahl einer bestimmten Rechtsfolge (Strafrahmen) besteht, sondern auch in der Wahl eines bestimmten Strafmaßes innerhalb dieses Strafrahmens. Wie man das optimale Strafmaß innerhalb eines Strafrahmens erhält, ist eine sehr wichtige Frage. Die Beantwortung dieser Frage wird allerdings auf später verschoben; erst im 4. Kapitel wollen wir uns mit dieser Problematik eingehend befassen und dabei das Modell von Haag und dessen Modifikation vorstellen.

§3. Verschiedene entscheidungstheoretische Modelle und die Frage ihrer Anwendbarkeit auf die Strafrechtspraxis

In diesem Paragraphen befassen wir uns mit der grundsätzlichen Frage, ob man eine Theorie der rationalen richterlichen Entscheidungen entwickeln kann, die im Rahmen einer bestimmten entscheidungstheoretischen Konzeption gewisse juristisch relevante Entscheidungssituationen adäquat wiedergibt, und wenn ja, welche entscheidungstheoretische Konzeption dafür am geeignetsten erscheint. Wie wir gesehen haben, hat man mit der Ausnahme des konservativen Entscheidungsmodells in allen Modellen es als selbstverständlich und ohne das weiter zu begründen angenommen, daß eine bestimmte Art von Entscheidungssituationen vorliegt, die durch das Vorkommen folgender Entitäten gekennzeichnet ist: Sachverhalte, Entscheidungen und Konsequenzen, die durch Sachverhalte und Entscheidungen bestimmt werden. Den Sachverhalten werden Wahrscheinlichkeiten, den Konsequenzen Nützlichkeiten zugeordnet, anhand derer man für jede mögliche Entscheidungsalternative ihren Erwartungsnutzen ableiten kann. Diejenige Entscheidungsalternative, die den höchsten Erwartungswert hat, soll von einem rationalen Individuum gewählt werden¹⁾.

Aus der Entscheidungstheorie dagegen weiß man, daß es mindestens drei verschiedene entscheidungstheoretische Konzeptionen gibt: die von Savage, Fishburn und Jeffrey. Die oben beschriebene Art der Entscheidungssituation wäre demnach ein Savage-Modell. Dies wollen wir nun eingehend untersuchen²⁾.

Vergegenwärtigen wir uns eine juristische Entscheidungssituation³⁾ Der rationale Richter X steht vor der Entscheidung, ob er Y, angeklagt einer bestimmten Straftat Z, verurteilen oder freisprechen soll. Falls Y verurteilt wird, kommt nur eine bestimmte Rechtsfolge (ein bestimmter Strafraum) in Betracht, die durch die Gesetzgebung (für diese Art von Straftaten) festgelegt ist. Nun muß der Richter, bevor er seine Entscheidung fällt, davon überzeugt sein, daß der Angeklagte die Straftat begangen bzw. daß er sie nicht begangen hat. Anstelle von "überzeugt sein" haben wir, aus den im I. Teil dieser Arbeit angeführten Gründen, eine schwächere Redeweise

1) Falls den Konsequenzen Disutilities zugeordnet werden, ist diejenige Entscheidungsalternative als die beste zu wählen, die den kleinsten Erwartungswert hat.

2) Derjenige, der nur am Aufbau der rationalen juristischen Entscheidungstheorie interessiert ist, kann die folgenden Erörterungen überschlagen und gleich zu der Lektüre der informellen Einführung in die juristische Entscheidungstheorie übergehen.

3) Wir setzen hier wiederum (vgl. §1) voraus, daß es sich um einen rationalen Richter handelt.

(und damit Voraussetzung) benutzt, nämlich die, daß der Richter wenigstens dazu imstande sein muß, die in Frage kommenden Sachverhalte wahrscheinlichkeitstheoretisch zu beurteilen.

Dem rationalen¹⁾ Richter muß außerdem klar sein, daß jede Entscheidung zusammen mit einem Sachverhalt einen bestimmten Zustand als Konsequenz zur Folge hat. So entspricht z.B. dem Sachverhalt "Y ist der Mörder" und der Entscheidung "Y wird zu einer lebenslangen Gefängnisstrafe verurteilt" die Konsequenz "Y ist der Mörder und Y wird zu einer lebenslangen Gefängnisstrafe verurteilt" oder abkürzend: "Der Mörder wird zu einer lebenslangen Strafe verurteilt". Wie in §1 setzen wir voraus, daß die Konsequenz eindeutig bestimmt ist, wenn die Entscheidung und der Sachverhalt festliegen.

Wie wir sehen, spielen in der obigen Entscheidungssituation drei unterschiedliche Arten von Entitäten eine Rolle: erstens: Handlungen (d.s. Entscheidungen der Art: "der Angeklagte wird freigesprochen", "der Angeklagte wird verurteilt", "der Angeklagte wird zur Strafe... verurteilt" usw.); zweitens: Sachverhalte (d.s. in der Vergangenheit ausgeführte strafbare Handlungen oder Straftaten, die nachgewiesen werden müssen); und drittens: die Konsequenzen, die durch Handlungen und Sachverhalte (juristisch) determiniert sind.

Daß wir auch im Zusammenhang mit Sachverhalten von Handlungen reden, soll nicht in der Weise mißverstanden werden, daß wir es in unserem entscheidungstheoretischen Konzept mit zwei Arten von Handlungen zu tun haben. Denn die in einer bestimmten Entscheidungssituation in Betracht kommenden strafbaren Handlungen haben den Charakter des "bereits Stattgefundenen". An ihnen ist nur noch interessant, daß durch sie ein bestimmter Zustand in der Welt verwirklicht wurde. Um Mißverständnisse jedoch zu vermeiden, werden wir anstatt des juristischen Begriffs "strafbare Handlung" im folgenden stets den Ausdruck "Sachverhalt" bzw. "Tatbestand" benutzen. Und natürlich sind Sachverhalte von den eigentlichen Handlungen, d.h. den richterlichen Entscheidungen, sowohl kausal wie auch probabilistisch unabhängig.

1) Ob dem konkreten Richter das auch tatsächlich klar ist, interessiert uns im Rahmen einer normativen Entscheidungstheorie nicht; vgl. dazu S.238f dieser Arbeit.

3.1 Die Konzeption von Savage¹⁾

Unter einem Savage-Modell²⁾, auch Grundmodell genannt, versteht man ein 7-Tupel $\langle X, \mathcal{U}, \mathcal{R}, H, P, u, E_u \rangle$, wobei gilt:

- (1) X ist das Entscheidungssubjekt bzw. der Handelnde;
- (2) $\mathcal{U}, \mathcal{R}, H$ sind nicht leer und endlich, $\mathcal{U}$ ist die Menge der möglichen Umstände (bei Savage sind es "Weltzustände" oder kurz "Zustände"), $\mathcal{R}$ ist die Menge der möglichen Konsequenzen (Resultate) und H eine Menge von Funktionen von $\mathcal{U}$ in $\mathcal{R}$, welche die dem Handelnden offenstehenden Handlungen repräsentieren³⁾
- (3) P ist ein Wahrscheinlichkeitsmaß auf $\text{Pot}(\mathcal{U})$ ⁴⁾; Wahrscheinlichkeiten werden den (handlungsunabhängigen) Ereignissen, d.h. den Elementen von $\text{Pot}(\mathcal{U})$ bzw. den Teilmengen von $\mathcal{U}$ ⁵⁾ zugeordnet; für jedes Ereignis $E_i \subset \mathcal{U}$ gilt $P(E_i) = \sum_{U_i \in E_i} P(U_i)$.

1) L.J.Savage, The Foundations of Statistics (1954).

2) Vgl. W.Spohn, Grundlagen der Entscheidungstheorie (1978), S.44f.

3) Im §1 haben wir die Konsequenzen konstruiert, indem wir eine zweistellige Funktion, definiert für Handlungen als ersten Argumentwerten und Umständen als zweiten Argumentwerten, angenommen haben, deren Funktionswerte die Resultate R_{ik} waren. Demgegenüber definiert Savage die Handlungen als Abbildungen der Menge der Umstände auf die Menge der Konsequenzen. Es sind somit einstellige Funktionen. Man kann leicht zeigen, daß dies eine andere Formulierung desselben Sachverhalts ist und daß die beiden Formalisierungen sich lediglich in unwesentlichen technischen Details unterscheiden. Läßt man nämlich in der ursprünglichen Formalisierung den ersten Index von R_{ik} für ein bestimmtes i fest und den zweiten Index k von 1 bis n laufen, so erhält man genau eine einstellige Funktion, welche die Handlung h_i repräsentiert und welche eine Abbildung der Klasse der Umstände auf die Menge der Resultate darstellt. Savage zieht es also vor (rein aus Zweckmäßigkeitsüberlegungen) jede der in Frage kommenden Handlungen durch eine eigene Funktion darzustellen (statt die Konsequenzenmatrix durch eine zweistellige Funktion zu repräsentieren). Die Matrix (R_{ik}) ist dann als durch m verschiedene einstellige Funktionen (mit $i = 1, \dots, m$) erzeugt aufzufassen. Vgl.dazu auch Stegmüller, Bd.IV, Teil B, S.291.

4) $\text{Pot}(\mathcal{U})$ bezeichnet die Potenzmenge von $\mathcal{U}$ (= die Menge aller Teilmengen von $\mathcal{U}$); vgl.S.10f dieser Arbeit; dort ist auch die formale Erläuterung zum Begriff "Möglichkeitsraum" (in unserer jetzigen Terminologie der "Raum der möglichen Umstände") zu finden.

5) Zum Begriff "Ereignis" vgl.S.11 dieser Arbeit. Die Elemente von $\mathcal{U}$ sind also elementare Ereignisse (zum Begriff "elementares Ereignis" siehe ebenda). Man kann Ereignisse, die in einer bestimmten Situation wahrscheinlichkeits-theoretisch beurteilt werden, ihrerseits in eine Menge $\mathcal{E}$ disjunkter und erschöpfender Ereignisse zusammenfassen; diese sollte nicht mit der Potenzmenge $\text{Pot}(\mathcal{U})$ verwechselt werden. Für $\mathcal{E}$ gilt lediglich $\mathcal{E} \subset \text{Pot}(\mathcal{U})$.

- (4) u ist eine Funktion von $\mathcal{R}$ in $\mathbb{R}$ ($\mathbb{R}$ ist die Menge der reellen Zahlen); u drückt die subjektiven Werte¹⁾ des Entscheidungssubjekts für die möglichen Konsequenzen aus;
- (5) E_u ist die auf H definierte Funktion, für die gilt: für alle $h \in H$
$$E_u(h) = \sum_{U_i \in \mathcal{U}} u(h(U_i)) \cdot P(U_i),$$

 E_u wird gewöhnlich als der erwartete subjektive Wert²⁾ der Handlung h bezeichnet.

Die mit Savage-Modellen verbundene normative Regel besagt: Wird X durch das Savage-Modell $\langle X, \mathcal{U}, \mathcal{R}, H, P, u, E_u \rangle$ charakterisiert, so soll X ein $h \in H$ wählen, für das $E_u(h)$ maximal ist.

Ein Savage-Modell ist dadurch charakterisiert, daß es einerseits zwischen drei Arten von Entitäten - Handlungen, Umständen und Konsequenzen - unterscheidet, andererseits verlangt, daß das Entscheidungssubjekt sich bei seiner probabilistischen Beurteilung auf eine Menge disjunkter und erschöpfender Ereignisse beschränkt und daß außerdem diese Ereignisse handlungsunabhängig sind. In formaler Hinsicht ist das Savage-Modell also das, was wir brauchen. Denn wie wir in dem eingangs erwähnten Beispiel festgestellt haben, muß der Richter - was durch die Rechtspraxis selbst bedingt ist - die in Frage kommenden Sachverhalte als solche (d.h. unabhängig von den möglichen Entscheidungen bzw. Konsequenzen) beurteilen und sie feststellen, bevor er seine Entscheidung trifft. Die Konsequenz seiner Entscheidung wird sowohl von dieser Entscheidung als auch vom Sachverhalt abhängen, jedoch weder mit der ersteren noch mit dem letzteren identisch sein³⁾

1) diese sind uns bereits unter der Bezeichnung "Nützlichkeiten bzw. Wünschbarkeiten" bekannt.

2) auf S.246f dieser Arbeit haben wir E_u als "erwarteten Nutzen" bezeichnet. Die oben verwendeten Ausdrücke "subjektiver Wert" bzw. "erwarteter subjektiver Wert" sind offensichtlich in dem Sinn neutraler, als sie mit keinen (überflüssigen) Assoziationen bei ihrer Interpretation belastet werden. Bei unserer nachfolgenden Darstellung eines rationalen juristischen Entscheidungsmodells werden wir wieder den alten Begriff "subjektiver (erwarteter) Nutzen" verwenden.

3) Wie allerdings das konservative Entscheidungsmodell (S.252 dieser Arbeit) zeigt, werden die Konsequenzen einer juristischen Entscheidung in der traditionellen Rechtspraxis nicht bzw. nicht explizit berücksichtigt. Da dies - wendet man die konservative Entscheidungsregel konsequent an - zu unsinnigen Ergebnissen führen

Daß diese Sachverhalte dabei erschöpfend, disjunkt und handlungsunabhängig sind, ist offensichtlich. Handlungsunabhängig sind sie schon deswegen, weil sie in der Vergangenheit liegen und von den gegenwärtigen Entscheidungen des Richters gar nicht beeinflußt werden können. Daß sie erschöpfend sind, liegt von vornherein im Interesse des Richters (denn der Richter muß alle seiner Ansicht nach in Frage kommenden Tatbestände überprüfen). Und daß sie disjunkt (d.h. sich wechselseitig ausschließend) sind, ist durch die Gesetzgebung sichergestellt, nach der nur ein Tatbestand in Frage kommt, d.h. es kann nicht etwa "Mord" und "Totschlag" gleichzeitig vorliegen, sondern entweder nur "Mord" oder nur "Totschlag"¹⁾

Das Savage-Modell scheint die juristische Entscheidungssituation noch in zwei weiteren Hinsichten adäquat wiederzugeben: Es werden darin nur Konsequenzen, d.h. weder den Umständen noch den Handlungen, von vornherein Nützlichkeiten zugeordnet²⁾ und es werden darin keine Handlungen probabilistisch beurteilt. Dies sei hier näher erläutert.

Ein Richter wird sicher zu vermeiden suchen, die Wahl seiner Entscheidung auch davon abhängig zu machen, daß er etwa von vorn-

kann (vgl. S.252 dieser Arbeit), muß man das konservative Modell durch ein entscheidungstheoretisches ersetzen, das auch die Wünschbarkeit aller möglichen Resultate berücksichtigt. Die Berücksichtigung der Konsequenzen und ihrer Nützlichkeiten neben den Sachverhalten stellt somit eine normative Forderung an eine konsistente und konsequente juristische Entscheidungspraxis dar.

1) Die Situation, in der man den sog. qualifizierten Tatbestand (wie z.B. "Raub") gegen den sog. Grundtatbestand (wie "Diebstahl") abwägt, sollte man nicht zu sehr problematisieren: Es wird zuerst der Grundtatbestand "Diebstahl" gegen "Nicht-Diebstahl" abgewogen, was eine erschöpfende Disjunktion bildet. Stellt man mit einer ausreichenden Wahrscheinlichkeit den Grundtatbestand fest, so wird das Vorhandensein zusätzlicher Merkmale, die den qualifizierten Tatbestand charakterisieren (wie z.B. "Waffengebrauch") gegen ihr Nichtvorliegen abgewogen, was wiederum eine erschöpfende Disjunktion darstellt. Überwiegen diese probabilistisch (bzw. sind sie mit einer ausreichenden Wahrscheinlichkeit festgestellt), so wird der Richter anstatt der ursprünglichen eine neue Entscheidungssituation konstruieren, der diesmal drei disjunkte und erschöpfende Ereignisse zugrunde liegen. Ausführlicher wird dies auf S.297ff dieser Arbeit behandelt.

2) Zwar erhalten im Modell auch Handlungen einen Nutzen, den sog. Erwartungsnutzen. Dieser wird allerdings den Handlungen nicht direkt zugeordnet, sondern erst aus den mit den Wahrscheinlichkeiten für die Umstände gewichteten Nützlichkeiten für die Konsequenzen abgeleitet.

herein eine besondere Vorliebe für bestimmte Arten von Strafen hat bzw. daß er bestimmte Strafen von vornherein ablehnt oder daß ihm ein bestimmter Sachverhalt, etwa, daß gerade dieser Angeklagte den Mord begangen hat, besonders wünschenswert erscheint (Sonst wäre der betreffende Richter für seinen Beruf ungeeignet). Während das additive Hinzufügen von (handlungsunabhängigen) Nützlichkeiten für die Umstände zu den Nützlichkeiten für die Konsequenzen die Präferenzordnung der Handlungen offensichtlich nicht verändert¹⁾, gilt das nicht in bezug auf die Nützlichkeiten für die Handlungen. Betrachten wir dazu das folgende Beispiel:

Der Entscheidende beurteile folgende Entscheidungssituation:

	schönes Wetter	schlechtes Wetter
h_1 : Fahren mit Zug	8 Std.	8 Std.
h_2 : Fliegen	4 Std.	14 Std.

Seine Nützlichkeits- und Wahrscheinlichkeitsmatrix hätten jeweils die Gestalt:

$$\begin{bmatrix} 0.5 & 0.5 \\ 1 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 0.7 & 0.3 \\ 0.7 & 0.3 \end{bmatrix}$$

Daraus berechnet das Entscheidungssubjekt die Erwartungswerte der Nutzen:

$$E_u(h_1) = 0.5 \cdot 0.5 + 0.5 \cdot 0.3 = 0.4$$

und

$$E_u(h_2) = 1 \cdot 0.7 + 0 \cdot 0.3 = 0.7.$$

Somit muß es sich für die zweite Handlung entscheiden. Nun entsinnt sich der Entscheidende, daß ihm das Fahren mit dem Zug an sich mehr

1) Man kann nämlich leicht folgende allgemeine Behauptungen für die handlungsunabhängige Nützlichkeitsmatrix beweisen: Seien $b_1, \dots, b_n$ verschiedene (handlungsunabhängige) Nützlichkeiten, die durch reelle Zahlen repräsentiert werden. Wenn wir b_1 zu den Werten der ersten Spalte der Nützlichkeitsmatrix addieren, b_2 zu denen der zweiten Spalte, usw., so erhalten wir wieder eine äquivalente Nützlichkeitsmatrix, denn bei der Bestimmung der subjektiven Erwartungswerte der Handlungen h_i wird jedesmal derselbe Wert $\sum_{j=1}^n b_j p_j$ addiert.

Diese Behauptung gilt allerdings nicht für die handlungsabhängigen Nützlichkeiten der Umstände. Denn in diesem Fall wird nicht nur zu jeder Spalte, sondern auch zu jeder Zeile ein neuer Wert addiert. Die Präferenzordnung der Handlungen kann sich somit ändern.

Vergnügen bereitet, sei es, weil er dabei die Landschaft genießen kann, oder sei es, weil er sich im Zug viel sicherer als im Flugzeug fühlt. Er bewertet deshalb das Fahren mit dem Zug, d.h. $u(h_1)$ mit 1 und das Fliegen, d.h. $u(h_2)$ mit 0.1. Die Erwartungswerte verändern sich dementsprechend zu:

$$E_u(h_1) = (0.5+1) \cdot 0.7 + (0.5+1) \cdot 0.3 = 1.5$$

und

$$E_u(h_2) = (1+0.1) \cdot 0.7 + (0+0.1) \cdot 0.3 = 0.8.$$

Der Entscheidende muß also, falls er auch die Tatsache berücksichtigt, wie wünschenswert ihm seine Handlungen sind (unabhängig von den in der gegebenen Situation zu beurteilenden Ereignissen und den möglichen Konsequenzen) seine Entscheidung evtl. modifizieren. Dasselbe gilt offensichtlich in Bezug auf die handlungsabhängigen Nützlichkeiten für die Umstände¹⁾ Für den Alltag mag das nichts besonderes darstellen, denn wir korrigieren fortwährend nicht nur unsere subjektiven Wahrscheinlichkeitsannahmen je nachdem, relativ auf welche Informationen hin sie beurteilt werden, sondern auch unsere Nützlichkeitsbewertungen von Ereignissen je nachdem, welcher Aspekt dabei berücksichtigt wird. Für die Rechtspraxis wäre aber eine solche Erweiterung auf die Entscheidungen bzw. Sachverhalte verheerend. Die Einschränkung, daß der Richter seine Entscheidung unabhängig davon macht, wie lieb (aus pädagogischen oder sonstigen Gründen) ihm diese oder jene Entscheidung von vornherein ist, bzw. wie wünschenswert ihm dieser oder jener Sachverhalt in Abhängigkeit davon erscheint, wie er sich entscheidet (etwa, daß der Angeklagte ein Mörder ist, während er die Bestrafung für Totschlag, minder schwerer Fall des Totschlags usw. wählt) ist aber im Savage-Modell trivialerweise erfüllt²⁾

Auch die Bewertung der möglichen Entscheidungen mit Wahrscheinlichkeiten ist anscheinend mit der juristischen Praxis unvereinbar: Denn eine solche Wahrscheinlichkeitsbewertung der Entscheidung

1) Vgl. Fn. 1 vorige Seite.

2) Werden allerdings die Konsequenzen mit den geordneten Paaren (h_i, u_k) , d.h. mit den Paaren, deren erstes Argument eine Handlung und deren zweites Argument ein Umstand ist, identifiziert - wie noch später in bezug auf die juristischen Konsequenzen zu erläutern sein wird - so läuft offensichtlich obige Restriktion darauf hinaus, daß nur den Konjunktionen $h_i \wedge u_k$, nicht dagegen ihren Bestandteilen Nützlichkeiten zugeordnet und keine (durch Handlungen) bedingten Nützlichkeiten (für Umstände) angenommen werden können.

würde mit dem Nachweis eines konkreten Sachverhalts nichts mehr zu tun haben; vielmehr würden dann auch solche Faktoren (vielleicht sogar ausschlaggebend) die Wahl der Entscheidung beeinflussen, die in der jeweiligen Entscheidungssituation überhaupt nicht vorkommen (etwa die Wahrscheinlichkeit, daß der Angeklagte bestraft wird aufgrund der richterlichen Erwartung, daß er bestraft wird). Juristisch gesehen wäre das unzulässig, denn die Lösung eines konkreten Rechtsfalls wäre dann entweder durch die Überbewertung des der Entscheidung entsprechenden Tatbestands (durch die Berücksichtigung seiner Wahrscheinlichkeit bei der probabilistischen Bewertung der Handlungen) oder durch irrelevante Mutmaßungen seitens des Richters (in Form probabilistischer Prognosen über die Entscheidungswahl) beeinflusst.

Was die Anwendung betrifft, so ist der Konzeption von Savage selbst wenig zu entnehmen. Ein aus purem Zufall Savage lesender Jurist wird kaum darauf kommen, daß Savage in seinem Buch gerade die Art von Situationen ins Auge faßt, mit denen man in der Rechtspraxis tagtäglich konfrontiert wird (Natürlich ist diese Art von Entscheidungssituationen nicht nur für die Rechtspraxis kennzeichnend und gewiß hat Savage nicht an die Rechtspraxis gedacht, als er seine Theorie aufstellte).

Damit aus dem Modell eine juristische Entscheidungstheorie entsteht, müssen also nicht nur die Grundbegriffe juristisch interpretiert, sondern auch die Forderungen, die an die Rechtspraxis gestellt sind, weitgehend berücksichtigt werden. Ob man dabei allen Forderungen gerecht wird, soll hier offen bleiben. Eines sollte man nicht vergessen: Wir haben es hier nicht mit einer empirischen Theorie zu tun, die die juristischen Verhältnisse so beschreibt, wie sie sind, bzw. wie sie gewöhnlich in der Rechtsliteratur (als tatsächlich bzw. erstrebenswert) dargestellt werden, sondern mit einer normativen (rationalen) Entscheidungstheorie, die eine Art Empfehlung an einen Präzision und Klarheit bevorzugenden Rechtsanwender gibt.

3.2. Die Konzeption von Fishburn¹⁾

Unter einem Fishburn-Modell²⁾ versteht man ein 6-Tupel $\langle X, C, H, P, u, E_u \rangle$ wobei gilt:

- (1) X ist das Entscheidungssubjekt oder Individuum.
- (2) C und H sind nicht leer und endlich, C ist die Menge der handlungsbedingten Konsequenzen, H die Menge der Handlungen.
- (3) P ist eine auf H definierte Funktion derart, daß für jedes $h \in H$ P_h ein Wahrscheinlichkeitsmaß auf $\text{Pot}(C)$ ist.
- (4) u ist eine Funktion von C in $\mathbb{R}$.
- (5) E_u ist die Funktion von H in $\mathbb{R}$, für die gilt:
für alle $h \in H$
$$E_u(h) = \sum_{c \in C} u(c) \cdot P_h(c).$$

Die normative Regel lautet: Wird X durch das Fishburn-Modell $\langle X, C, P, u, E_u \rangle$ charakterisiert, so soll X eine Handlung $h \in H$ wählen, für die $E_u(h)$ maximal ist.

Wie wir sehen, verschwinden aus der Konzeption Fishburns die möglichen Umstände und mit ihnen die möglichen Ereignisse als gesonderte Elemente einer Entscheidungssituation völlig; außerdem werden handlungsbedingte subjektive Wahrscheinlichkeiten für die Konsequenzen zugelassen³⁾ Wie ist das zu erklären? Man kann Fishburns Konzeption als eine Reaktion auf das Modell von Savage verstehen, das nicht allgemein genug ist: Bestimmte Entscheidungssituationen, insbesondere solche, in denen es nicht auf die möglichen Umstände ankommt, können in ihm nicht bzw. nicht direkt dargestellt werden, wie etwa im folgenden Beispiel:

1) P.C.Fishburn, Decision and Value Theory (1964).

2) Vgl.dazu W.Spohn, S.67f.

3) Zwar kommen die möglichen Umstände als unabhängige Komponenten einer Entscheidungssituation nicht vor, handlungsunabhängige Ereignisse können jedoch in einem Fishburn-Modell nach wie vor eine Rolle spielen, nämlich als Bestandteile der Konsequenzen. Da Savage-Konsequenzen und Fishburn-Konsequenzen voneinander verschieden sein können, wird dieser Tatsache im Modell durch die Verwendung unterschiedlicher Symbole "R" für Savage-Konsequenzen (entsprechend "X" für die Menge des Savage-Konsequenzen) und "c" für Fishburn-Konsequenzen (entsprechend "C" für die Menge der Fishburn-Konsequenzen) Rechnung getragen.

X sitze in einem Restaurant und müsse sich für eines der beiden angebotenen Tagesmenüs entscheiden. Es sei dabei angenommen, daß X unbedingt ein Tagesmenü bestellen will und daß es ihm außerdem nur auf den Geschmack der Speise ankommt. X wird in dieser Situation offensichtlich dasjenige Tagesmenü wählen, bei dem er die Aussicht, daß es ihm auch schmeckt, höher einschätzt.

Die obige Entscheidungssituation kann man so auffassen, daß die X offenstehenden Handlungen in der Wahl des ersten bzw. des zweiten Tagesmenüs bestehen und als mögliche Konsequenzen "Schmecken" und "Nichtschmecken" ansehen. Allerdings kommt es hier offensichtlich bei der Wahl einer Handlung nur auf die subjektive Bewertung der Konsequenzen an. Irgendwelche Umstände sind dagegen in der obigen Entscheidungssituation anscheinend ohne Relevanz: Entweder sie bieten sich nicht an oder sie sind nicht (ohne weiteres) anzugeben¹⁾.

Es ist klar, daß ein Fishburn-Modell nicht geeignet ist, die juristische Entscheidungssituation zu repräsentieren: Denn in dieser kommt es primär gerade darauf an, die Umstände (Sachverhalte) gesondert auszuzeichnen und ausschließlich sie wahrscheinlichkeits-theoretisch zu beurteilen. Wir können somit die Umstände nicht verschwinden lassen, indem wir sie etwa innerhalb der Konsequenzen als deren Bestandteile betrachten²⁾³⁾

Nur die Handlungen und die Konsequenzen zu betrachten, wäre auch noch aus dem folgenden Grund wenig befriedigend: Im eingangs besprochenen Beispiel hätten wir dann 2 Handlungen ("Verurteilung"/

1) Wir wollen die Darstellung eines Fishburn-Modells an dieser Stelle mit dem Hinweis abbrechen, daß die Analyse eines solchen viel zu tiefgreifend ist, als daß sie uns hier möglich wäre. Vgl. dazu W.Spohn (S.48ff und 66ff), der sich mit dieser Problematik im Rahmen einer Dissertation eingehend befaßt. Wie Spohn (S.68) zeigt, ist das Fishburn-Modell eine Verallgemeinerung des Savage-Modells, d.h. man kann jede Entscheidungssituation à la Savage in eine Entscheidungssituation à la Fishburn übersetzen, indem man die Menge der Konsequenzen dem Kreuz-Produkt aus den ursprünglichen Mengen der Umstände und der Konsequenzen gleich setzt und die handlungsbedingten Wahrscheinlichkeiten für die neuen Konsequenzen mit den alten Wahrscheinlichkeiten für die Umstände identifiziert. Auch das umgekehrte Verfahren - die Übersetzung einer Fishburn- in eine Savage-Entscheidungssituation - funktioniert, allerdings nicht ohne gewisse Restriktionen und sprachliche Verdrehungen.

2) Vgl.S.277 Fn.3 dieser Arbeit.

3) Der Ausdruck "nicht geeignet" im obigen Abschnitt ist nicht so zu verstehen, daß eine juristische Entscheidungssituation im Fishburn-Modell nicht rekonstruierbar wäre. Das dürfte nach Spohn (siehe auch die Anmerkung in Fn.1 oben) prinzipiell möglich sein. Doch eine solche Rekonstruktion wäre im juristischen Fall ein unnötiger und intuitiv gesehen wenig plausibler Vorgang.

"Freispruch") und vier Konsequenzen ("Verurteilung eines Mörders", "Verurteilung eines Nicht-Mörders", "Freispruch eines Mörders", "Freispruch eines Nicht-Mörders"). Diesen Konsequenzen müßte man dann handlungsbedingte Wahrscheinlichkeiten zuordnen. Die letzteren brauchen aber nicht mit den Wahrscheinlichkeiten, daß "Y der Mörder ist" bzw. daß "Y nicht der Mörder ist" identisch zu sein. Im Unterschied zu einem Savage-Modell, das so konstruiert ist, daß Wahrscheinlichkeiten nur den Umständen zugeordnet werden, was die Gefahr bannt, daß irgendwelche sonstigen probabilistischen Überlegungen (etwa in bezug auf Konsequenzen) in das Modell Eingang finden, ist dies im Modell von Fishburn nicht garantiert: Da dort den Konsequenzen Wahrscheinlichkeiten zugeordnet werden, können an deren Feststellung auch Gesichtspunkte (etwa statistischer Art) maßgebend beteiligt sein, die mit dem konkreten Sachverhalt nichts zu tun haben.

Terminologisch gesehen, sind wir außerdem nicht an der Wahrscheinlichkeit interessiert, daß der Mörder verurteilt wird, sondern lediglich an derjenigen, daß Y der Mörder ist - und dies auch dann, wenn beide Wahrscheinlichkeiten identisch wären. Es macht eben nicht nur einen inhaltlichen, sondern auch einen sprachlichen Unterschied, ob man von der Wahrscheinlichkeit der Konsequenz "Der Mörder wird verurteilt" oder von der Wahrscheinlichkeit des Sachverhalts "Der Angeklagte ist der Mörder" redet.

3.3. Die Konzeption von Jeffrey¹⁾

Unter einem Jeffrey-Modell²⁾ versteht man ein 5-Tupel $\langle X, W, H, P, u \rangle$ wobei gilt:

- 1) X ist das Entscheidungssubjekt oder das Individuum.
- 2) W ist eine nicht leere, endliche Menge.
- 3) H ist einer Zerlegung von W .
- 4) P ist ein Wahrscheinlichkeitsmaß auf $\text{Pot}(W)$; die Elemente von $P(W)$ sind Propositionen.
- 5) u ist eine Funktion von $\text{Pot}(W) \setminus \{\emptyset\}$ in $\mathbb{R}$, für die gilt:

Für $A, B \subset W$ mit $P(A \cap B) = 0$ und $P(A \cup B) \neq 0$ ist

$$\begin{aligned} u(A \cup B) &= \frac{u(A) \cdot P(A) + u(B) \cdot P(B)}{P(A) + P(B)} \\ &= u(A) \cdot P(A/A \cup B) + u(B) \cdot P(B/A \cup B). \end{aligned}$$

Die normative Regel lautet: Wird X durch das Jeffrey-Modell $\langle X, W, H, P, u \rangle$ charakterisiert, so soll X ein $A \in H$ wählen, für das $u(A)$ maximal ist.

Wie wir sehen, werden die unterschiedlichen Entitäten, wie Handlungen, Konsequenzen und Umstände, welche in den Modellen von Savage und Fishburn (bei letzterem nur Handlungen und Konsequenzen) auftraten, im Jeffrey-Modell durch eine einheitliche Art von Entitäten - Propositionen, d.h. solcher Objekte, auf die sich die Daß-Sätze in Ausdrücken der Form " X glaubt, daß..." bzw. " X wünscht, daß..." beziehen - ersetzt. Damit wird die Trennung von Wünschens- und Glaubensbereichen hinfällig: Ein und dieselbe Proposition kann im Definitionsbereich sowohl der subjektiven Wahrscheinlichkeits- als auch der subjektiven Wertefunktion auftreten³⁾. Da Handlungen auch Propositionen sind, kann man ihnen jetzt ebenfalls Wahrscheinlichkeitswerte und (nichtabgeleitete) Nutzenswerte zuschreiben. Wie verheerend das für die Rechtspraxis sein kann (denn ob jemand verurteilt oder nicht verurteilt wird, hängt nun nicht allein davon ab, ob er der Täter ist oder nicht, sondern auch von irgendwelchen sonsti-

1) R.C.Jeffrey, Logik der Entscheidungen (1967).

2) Vgl. W.Spohn, S.77.

3) Formal wird diese "einheitliche Art von Entitäten" bzw. "Propositionen" im Modell durch die Verwendung eines einzigen Symbols " W " dargestellt. Da der Unterschied zwischen dem Nutzen und dem erwarteten Nutzen wegfällt, bringt man das im Modell dadurch zum Ausdruck, daß man das gleiche Symbol für beide verwendet.

gen Faktoren), läßt sich hier nur erahnen¹⁾

Aus dem oben genannten Grund können wir die Konzeption von Jeffrey nicht auf die Rechtspraxis anwenden.

Zusammenfassung und Ausblick

Die Untersuchung der drei wichtigsten entscheidungstheoretischen Modelle hat uns gezeigt, daß das Modell von Savage den geeigneten Rahmen für den Aufbau einer rationalen juristischen Entscheidungstheorie bildet, denn darin werden bestimmte, eine juristische Entscheidungssituation kennzeichnende Charakteristika formal erfüllt, insbesondere die Auszeichnung einer gesonderten Menge von handlungsunabhängigen Umständen, welche ausschließlich probabilistisch beurteilt werden. Außerdem konnte die intuitive Auffassung, die den von der juristischen Seite vorgeschlagenen Entscheidungsmodellen zugrunde liegt, und in denen man unwissentlich, jedoch selbstverständlich von einer Savage-Entscheidungssituation ausgeht, bestätigt werden.

Das Savage-Modell ist für juristische Zwecke allerdings noch zu allgemein. Vieles, was wir am Ende von §2 von einem adäquaten juristischen Entscheidungsmodell verlangten, wird von ihm nicht erfüllt, z.B. die Wahl einer juristisch konformen Entscheidung, falls mehrere Entscheidungsmöglichkeiten den gleichen maximalen Erwartungswert erhalten, oder die Verhinderung dessen, daß eine Entscheidung entsprechend dem unwahrscheinlichsten bzw. unwahrscheinlicheren Sachverhalt als die beste abgeleitet wird. Wir müssen somit an das Entscheidungsmodell zusätzliche Restriktionen (in Form von Postulaten) stellen, falls wir ein nicht nur entscheidungstheoretisch konsistentes, sondern auch juristisch adäquates Modell formulieren wollen.

Darüber hinaus müssen wir entscheidungstheoretische Begriffe juristisch interpretieren und uns über den Begriff des juristischen Nutzens einige Gedanken machen, wenn wir außerdem eine rationale juristische Entscheidungstheorie aufstellen wollen. Mit diesen und ähnlichen Fragen werden wir uns im nächsten Kapitel näher befassen.

1) Zur Kritik an der Zuordnung von Wahrscheinlichkeiten und (nicht abgeleiteten) Nützlichkeiten zu juristischen Handlungen und an der Zuordnung von Nützlichkeiten zu juristischen Sachverhalten vgl. S.272ff dieser Arbeit.

2. Kapitel. Rationale juristische Entscheidungsbildung

Wie bereits festgestellt, handelt es sich bei juristischen Entscheidungen bzgl. dessen, ob der Angeklagte verurteilt oder freigesprochen werden soll, um solche unter Risiko, denn die Sachverhalte, die solchen Entscheidungen zugrunde liegen, werden i.d.R. nicht mit Sicherheit feststellbar sein. Was aber bedeutet "Verurteilung eines Angeklagten" im bislang von uns benutzten Sinne? - Offensichtlich, daß lediglich ein gesetzlich vorgeschriebener Strafraum, z.B. "die Strafe nach §211 des StGB", gewählt wird. Damit ist natürlich die juristische Entscheidungsbildung noch längst nicht zu Ende. So sind wir bis jetzt weder darauf eingegangen, daß einzelne juristische Entscheidungen miteinander koordiniert werden müssen, falls mehrere Richter am Entscheidungsprozeß beteiligt sind, noch haben wir die Problematik bzgl. der Konkretisierung des Strafraums, d.h. der Wahl eines Strafmaßes innerhalb des Strafraums, erläutert. (Wie bereits erwähnt, ist die Wahl eines Strafmaßes, falls man sich bereits für einen bestimmten Strafraum entschieden hat, nicht mehr eine solche unter Risiko, sondern eine unter Sicherheit, deswegen muß die daraus entstehende Problematik anders als mit den bisher angegebenen Methoden angegangen werden). Dies können wir allerdings erst dann nachholen, wenn wir den Rahmen für die rationale juristische Entscheidungsbildung abgesteckt haben.

In diesem Kapitel geht es uns vor allem darum, diesen Rahmen - in Gestalt einer rationalen Theorie juristischer Einzelentscheidungen - zu explizieren. Das wichtigste Ziel, das von uns dabei verfolgt wird, ist, das Grundmodell (Savage-Modell), welches sich als zu allgemein für juristische Zwecke erwiesen hat, so einzuschränken, daß daraus ein juristisches Entscheidungsmodell entsteht. Allerdings wird dieses ein sog. minimales Entscheidungsmodell sein, das zwar bestimmte unbefriedigende, aus dem Grundmodell sich ergebende Folgen verhindern kann, gewisse juristische Forderungen aber, wie etwa die, daß die Entscheidung dann ableitbar sein soll, wenn der entsprechende Sachverhalt "mit an Sicherheit grenzender Wahrscheinlichkeit"¹⁾ festgestellt ist, nicht erfüllt (wir werden übrigens formal zeigen, daß diese Forderung an das Modell sogar unsinnige Einschränkungen bzgl. der Nutzenzuordnung zu den Konsequenzen impliziert).

Wir beginnen mit der Präzisierung der juristischen Entscheidungssituation (§1). Dazu gehört u.a., daß wir entscheidungstheoretische Begriffe juristisch (um-) interpretieren.

1) falls darunter die von manchen Juristen vorgeschlagene Wahrscheinlichkeit von 0.99... verstanden wird. Zu dieser Problematik vgl. S.58 Fn.1 dieser Arbeit.

Sl. Die juristische Entscheidungssituation

Wir wollen hier die im ersten Kapitel eingeführten entscheidungstheoretischen Begriffe auf eine juristische Entscheidungssituation anwenden. Es geht uns also primär darum, die Grundentitäten, d.h. die möglichen Umstände, die möglichen Handlungen und die möglichen Konsequenzen, als juristische Begriffe zu begreifen und als solche zu klären.

Außerdem wollen wir uns in diesem Paragraphen mit der Problematik des sog. "qualifizierten Tatbestands" befassen. Ein qualifizierter Tatbestand entsteht aus einem "Grundtatbestand", wenn zusätzliche Merkmale zu dem letzteren hinzukommen. Um gleichzeitig die qualifizierten und die Grund-Tatbestände beurteilen zu können (wie das oft in der Rechtspraxis der Fall ist), machen wir von einer formalen Konstruktion Gebrauch. Es werden dabei Grundtatbestände mit sog. "elementaren Sachverhalten" und qualifizierte Tatbestände mit bestimmten Teilmengen der Sachverhaltsmenge identifiziert.

1.1 Intuitive Schilderung des rationalen juristischen Entscheidungsmodells

Aus dem bisherigen (1. Kapitel) haben wir erfahren, daß eine rationale Person (das Entscheidungssubjekt X) sowohl die möglichen Umstände, die möglichen Handlungen, die in der Konsequenzenmatrix erhaltenen Resultate und deren Nützlichkeiten, als auch die Wahrscheinlichkeiten der möglichen Umstände kennen muß. Mögliche Umstände, mögliche Handlungen, Resultate, Nützlichkeiten und Wahrscheinlichkeiten gehören somit zu den Daten, auf deren Grundlage rationale Erwägungen über die richtige Entscheidung anzustellen sind.

Wir wollen nun eine juristische Entscheidungssituation näher studieren: In einem Strafprozeß muß das Gericht (der Richter bzw. mehrere Richter) ein Urteil (eine Entscheidung) fällen. Das Urteil besteht entweder im Freispruch oder in der Verurteilung. Im Falle der Verurteilung unterscheiden wir zwischen einer Strafe (Gefängnisstrafe, Geldstrafe, zur Bewährung ausgesetzte Gefängnisstrafe etc., hinzu kann evtl. auch eine Nebenstrafe, z.B. Fahrverbot für eine bestimmte Zeit, kommen) und einer Maßregel (z.B. Unterbringung in einer sozialtherapeutischen Anstalt oder Führerscheinentzug) wobei wir noch weiter differenzieren müssen: letztlich kommt es bei der Entscheidung darauf an, die Dauer der Strafe, die Höhe der Geldstrafe usw. zu konkretisieren. Die Strafe bzw. die Maßregel setzt im herkömmlichen Sinne voraus, daß sowohl der zu sanktionierende Sachverhalt (der sog. objektive Tatbestand) wie auch die Schuld bzw. Schuldfähigkeit des Angeklagten (der subjektive Tatbestand)¹⁾ nachgewiesen

1) Juristisch gesprochen ist Schuld - darunter versteht man Vorsatz bzw. Fahrlässigkeit des zu sanktionierenden Sachverhalts - Voraus-

sind, d.h. man muß zuerst nachweisen, daß der Angeklagte z.B. die ihm zur Last gelegte Mordtat begangen hat und daß diese Tat ihm auch persönlich im Sinne der Verantwortlichkeit vorzuwerfen ist, bevor man ihn nach der deutschen Rechtsprechung zu lebenslanger Freiheitsstrafe verurteilen darf.

Wir nehmen an, daß es sich bei unserem Richter (kurz "X" genannt) um eine rationale Person handelt. Ein rationaler Richter ist also in unseren Augen eine Person, die relativ zu den ihr zur Verfügung stehenden Daten möglichst richtig (d.h. konsequent und konsistent) entscheiden will. Oben haben wir gesagt, daß zu diesen Daten u.a. auch die Konsequenzen mit ihren Nützlichkeiten gehören. In der herkömmlichen juristischen Entscheidungspraxis werden diese (wie bereits erwähnt) allerdings nicht explizit berücksichtigt¹⁾. Dies ist offensichtlich höchst unbefriedigend, denn die Folgen einer juristischen Entscheidung sind sowohl für den Angeklagten wie auch für die Gesellschaft von großer (oft entscheidender) Bedeutung. Müßte unser Richter z.B. im Alltag entscheiden, so würden wir ihn als nicht rational bezeichnen, falls er dies täte, ohne die Konsequenzen seiner Entscheidungen zu berücksichtigen. Die Strafrechtspraxis, in der es nicht um die Bestrafung an sich geht, sondern darum, durch die Strafe etwas zu erreichen bzw. zu verhindern, muß also so angelegt werden, daß im strafrechtlichen Entscheidungsprozeß auch Konsequenzen und deren Nützlichkeiten explizit berücksichtigt werden.

Die möglichen Umstände bzw. die möglichen Sachverhalte (z.B. "Mord"/"kein Mord")²⁾ werden wir in Anlehnung an den juristischen Sprach-

setzung der Strafbarkeit. Schuldfähigkeit dagegen ist Voraussetzung für die strafrechtliche Verantwortlichkeit (sie fehlt z.B. bei den zur Tatzeit noch nicht 14jährigen (§19 StGB), bei Geistes- bzw. bestimmten psychisch Kranken).

1) Vermutlich findet die Berücksichtigung der Konsequenzen auf Umwegen doch Eingang in den juristischen Entscheidungsprozeß, nämlich über verschiedene Theorien, wie z.B. Resozialisierungstheorie, Vergeltungstheorie und dergl., denn es macht einen Unterschied, ob man jemanden unter dem Aspekt der Resozialisierung verurteilt oder etwa dem der Vergeltung. Indem man bestimmte Zielsetzungen und -vorstellungen die Entscheidung mitkonstituieren läßt, sucht man offenbar einen Zustand (in unserer Terminologie: "Konsequenz") als wünschenswert zu erreichen bzw. als unerwünscht zu verhindern.

2) Dabei sind "Mord", "kein Mord" etc. als Abkürzungen zu lesen für: "Der Angeklagte ist der Mörder" oder "Y hat die Mordtat begangen", "Y ist kein Mörder" oder "Der Angeklagte hat den Mord nicht begangen" etc..

gebrauch "Tatbestände" nennen¹⁾, wobei wir einfachheitshalber "Mord", "Diebstahl", "Raub" etc. als positive Tatbestände und "kein Mord", "kein Diebstahl", "kein Raub" etc. als negative Tatbestände bezeichnen²⁾.

Die möglichen Handlungen sind nach den obigen Erläuterungen entweder "Freispruch" ("keine Strafe") oder "Verurteilung" ("Strafe", "Strafe nach §... des StGB", "Strafe wegen ...", "Maßregel ..." etc.)³⁾. Wir werden Strafen und Maßregeln in Anlehnung an Haag⁴⁾ unter dem Oberbegriff "Maßnahmen" zusammenfassen, wobei darunter zur Vereinfachung der Darstellung auch der Freispruch ("keine Strafe") zu subsumieren ist. Bei den Maßnahmen handelt es sich lediglich um gesetzliche Strafraumen, nicht dagegen um konkretisierte Strafen.

Konsequenzen werden mit den durch Tatbestände und Maßnahmen bedingten Ereignissen⁵⁾ identifiziert. Die Konsequenzen sind z.B. "Verurteilung eines Mörders", "Verurteilung eines Nicht-Mörders", "Strafe nach §... für Diebstahl" etc.. Dies sind wiederum Abkürzungen für bestimmte Sachverhalte. So besagt die Konsequenz "Strafe nach §242 StGB für Diebstahl" nichts anderes als "Y (der Angeklagte) hat den Diebstahl begangen und Y wird nach §242 StGB wegen Diebstahls verurteilt". Die so verstandenen Konsequenzen sind also ebensowenig

1) Natürlich gehört (wir sprachen bereits darüber auf S.55 Fn.2 dieser Arbeit) zum "Tatbestand" mehr als seine empirische Feststellung. So klammern wir aus unserem entscheidungstheoretischen Konzept die normative Begründung (etwa in Form einer Begriffsanalyse anhand von Kommentaren und Entscheidungssammlungen) aus zwei Gründen aus: Erstens tangiert die normative Begründung eines Tatbestands seine entscheidungstheoretische keineswegs, und zweitens bin ich der Ansicht, daß die normative Begründung von einem dafür speziell ausgebildeten Juristen besser bewältigt werden kann.

2) Im juristischen Sprachgebrauch kennt man natürlich nur "Mord", "Diebstahl" usw. als eigentliche Tatbestände.

3) Dies soll wiederum als Abkürzung verstanden werden. So bedeutet "Strafe" nichts anderes als "Der Angeklagte wird zu einer Strafe verurteilt".

4) Vgl. Haag, Rationale Strafzumessung (1971).

5) Die Identifizierung der Konsequenzen mit Ereignissen soll keineswegs so verstanden werden, daß wir von der Savage-Konzeption abweichen wollen, sondern nur verdeutlichen, daß Konsequenzen auch bestimmte Tatsachen in der Welt darstellen. Probabilistisch werden nach wie vor nur (handlungsunabhängige) Sachverhalte beurteilt. Im Savage-Modell waren Ereignisse Teilmengen von $Pot(\Lambda)$. Wir setzten dabei voraus, daß es sich bei diesen um handlungsunabhängige Ereignisse handelt, vgl. S.271ff dieser Arbeit.

Maßnahmen, wie die Ereignisse "Fahren mit dem Zug bei schönem Wetter", "Fliegen bei Sturm" Handlungen sind.

Wenn im folgenden von der Wahrscheinlichkeit des Tatbestands die Rede sein wird, so meinen wir damit zunächst die Wahrscheinlichkeit des Sachverhalts (d.h. des objektiven Tatbestands): Erst später im Laufe des Aufbaus der Theorie (und damit der Annäherung an die Rechtspraxis) werden wir die Tatsache berücksichtigen, daß man juristische Entscheidungen an den Gesamttatbestand (objektiver Tatbestand + subjektiver Tatbestand (Schuld bzw. Schuldfähigkeit)) anknüpft.

Im 1. Teil haben wir festgestellt, daß die probabilistische Beurteilung eines juristischen Ereignisses eine sog. bedingte Wahrscheinlichkeitsbewertung darstellt. Wir werden uns allerdings in diesem Teil aus Einfachheitsgründen der Ausdrucksweise der absoluten Wahrscheinlichkeiten bedienen, wohlwissend, daß es sich dabei in Wirklichkeit um bedingte Wahrscheinlichkeiten handelt. Wenn also im folgenden von der Wahrscheinlichkeit eines Sachverhalts die Rede sein wird, so ist das als bedingte Wahrscheinlichkeit zu verstehen, z.B. ist die Wahrscheinlichkeit für den Mord nichts anderes als die Wahrscheinlichkeit dafür, daß der Angeklagte den Mord begangen hat, relativ zu irgendwelchen (für diesen Sachverhalt relevanten) Informationen.

Bevor wir zur Anwendung übergehen, müssen wir uns darüber im klaren sein, daß der Entscheidende mit der Wahl seiner Entscheidung bestimmte Ziele verfolgt. Seine Zielvorstellungen werden im Entscheidungsmodell selbst nicht direkt berücksichtigt¹⁾. Sie finden jedoch ihren Niederschlag bei der subjektiven Bewertung der Konsequenzen seitens des Entscheidenden. Es ist nicht immer leicht, die Ziele anzugeben, welche einer konkreten Entscheidung zugrunde liegen. Hinsichtlich der Rechtspraxis sind wir jedoch im Vorteil: Strafrechtliche Ziele sind nichts anderes als Vorstellungen darüber, was die Strafe eigentlich bewirken soll, sei es in bezug auf den Täter oder in bezug auf die Öffentlichkeit.

Man kann die Strafe unter verschiedenen Aspekten sehen, dem der Vergeltung, dem der Abschreckung, dem des Ausgleichs des Übels, dem der Vorbeugung zukünftiger Straftaten seitens des Täters usw. Alle diese Vorstellungen werden in der Rechtswissenschaft durch gängige Theorien repräsentiert. So befaßt sich die Vergeltungstheorie mit der Strafe unter dem Gesichtspunkt der Vergeltung, die generalpräventive Theorie unter dem Gesichtspunkt der Allgemeinabschreckung, die Resozialisierungstheorie unter dem Gesichtspunkt der Wiedereingliederung des Täters in die Gesellschaft etc. Wir wollen

1) Obwohl sie natürlich bei der Spezifizierung dessen, was man in einer konkreten Entscheidungssituation für die möglichen Konsequenzen halten soll, eine Rolle spielen.

hier nicht auf alle diese Theorien eingehen, sondern unter den Zielen zunächst bestimmte oder unbestimmte Vorstellungen von Recht, Gerechtigkeit usw. verstehen, die demjenigen, der sie seinen Entscheidungen zugrunde legt, nicht einmal bewußt sein müssen; es sind dann eben seine kognitiven und normativen Grundlagen, die die Welt für ihn so konstituieren, wie er sie wahrhaben will. Jedoch können wir durch die Entscheidungstheorie gerade diese persönlichen Dispositionen aufdecken, was hinsichtlich der Frage nach den tatsächlichen Entscheidungsgründen sehr wichtig erscheint. Kurz: wir gehen zunächst von gewissen, juristisch relevanten und intuitiv plausiblen Zielvorstellungen aus. Erst später (§2) werden wir auf die mit der Rechtspraxis konsistenten Zielvorstellungen näher eingehen und die Begründung liefern für die Voranstellung der Gerechtigkeitseinsicht, wonach der Zweck der Strafe in der Wiederherstellung des sozialen Gleichgewichts bzw. des juristischen Status Quo, oder einfacher: im gerechten (tatangemessenen) Ausgleich des Unrechts zu sehen ist.

Mit diesem Begriffsmaterial wollen wir analog wie im 1. Kapitel (§1) vorgehen und die drei Matrizen: Konsequenzen-, Nützlichkeits- und Wahrscheinlichkeitsmatrix konstruieren.

Die Konsequenzenmatrix

Angenommen, unser Richter X muß zwischen zwei Maßnahmen: "Freispruch" ("keine Strafe") und "Verurteilung" ("Strafe") entscheiden. Dies sind also die beiden Handlungen, zwischen denen X wählen muß. Die möglichen Sachverhalte (Tatbestände) seien "Diebstahl" und "kein Diebstahl". Falls X sich für die Strafe entscheidet, wird die Konsequenz für beide Sachverhalte in der Bestrafung des Angeklagten bestehen (wobei laut §242 StGB auf "Diebstahl" evtl. eine Freiheitsstrafe bis zu fünf Jahren steht). Falls X sich für den Freispruch entscheidet, ist die Konsequenz in beiden Fällen die Nichtbestrafung des Angeklagten.

Wir merken bereits, daß die Entscheidung von X für ihn persönlich nichts ändert¹⁾, dagegen sehr viel für den Angeklagten. Auch die

1) Es sei denn, dadurch werden bestimmte Veränderungen in seinem Gemütsleben bewirkt; wobei anzumerken ist, daß ein gutes Gewissen bzw. Gewissensbisse manchmal gute psychologische Hilfen darstellen können, es aber nicht unbedingt müssen.

Gesellschaft hat ein großes Interesse daran, daß der Richter möglichst wenig Fehler macht. Kurz: der Richter darf nicht bei der Entscheidung seine persönlichen Interessen und Vorlieben in den Vordergrund stellen, sondern er muß versuchen, den Angeklagten und die Tat möglichst objektiv und neutral einzuschätzen¹⁾. Seine Entscheidung ist nichtsdestoweniger subjektiv, da er sowohl die möglichen Tatbestände als auch die möglichen Maßnahmen und ihre Konsequenzen subjektiv bewerten muß.

Die folgende Tabelle liefert einen Überblick über sämtliche Möglichkeiten²⁾:

	Diebstahl	kein Diebstahl
Strafe (nach §242)	Strafe für einen Dieb	Strafe für einen Nicht-Dieb
keine Strafe	keine Strafe für einen Dieb	keine Strafe für einen Nicht-Dieb

Ein anderes Beispiel:

Angenommen, es ist sicher, daß der Angeklagte das Opfer getötet hat. Nun will man wissen, ob es sich dabei um Mord oder Totschlag³⁾ handelt. Die möglichen Tatbestände sind somit: "Mord" und "Totschlag", wobei laut §211 StGB auf "Mord" lebenslänglich, laut §212 StGB auf "Totschlag" eine Freiheitsstrafe von nicht unter fünf Jahren folgt. Die Entscheidungssituation sieht wie folgt aus:

1) D.h. einerseits unabhängig von Faktoren, die mit der Tat nichts zu tun haben, andererseits ohne daß relevante Informationen unter den Tisch gekehrt werden. Dazu gehört auch, daß man möglichst viele Informationen gewinnt.

2) In den folgenden Konsequenzenmatrizen wird unter der entsprechenden Strafe immer die "tatschuldangemessene" Strafe, d.h. die durch die Gesetzgebung für diese Art von Straftaten vorgesehene Bestrafung verstanden (gewöhnlich durch den Strafraum angegeben).

Anstatt "Bestrafung des Täters nach §..." und dergleichen empfiehlt sich die folgende ein wenig hölzerne Schreibweise für die Konsequenzen, denn dadurch wird deutlicher, von welchen Sachverhalten und Maßnahmen die Konsequenzen jeweils abhängen.

3) "Mord" $\hat{=}$ "Tötung + zusätzliche Eigenschaften", "Totschlag" $\hat{=}$ "Tötung + Fehlen zusätzlicher Eigenschaften". Beachte: Da die Wahrscheinlichkeit für die Tötung gleich 1 ist, gibt es in der obigen Entscheidungssituation anstatt vier Ereignissen nur zwei. Zur Problematik des sog. qualifizierten Tatbestands vgl. S.297ff dieser Arbeit.

	Mord	Totschlag
lebenslange Strafe	lebenslänglich für Mord	lebenslänglich für Totschlag
nicht unter 5 Jahren	nicht unter 5 Jahren für Mord	nicht unter 5 Jahren für Totschlag

Aus diesen beiden Beispielen können wir analog dem Beispiel im ersten Kapitel, S.243 das allgemeine Schema abstrahieren, wobei wir die folgende geringfügige Modifikation berücksichtigen; Anstatt m Handlungen $h_1, \dots, h_m$ sind es diesmal r Maßnahmen (kurz: $M_1, \dots, M_r$ genannt); n mögliche Umstände $U_1, \dots, U_n$ werden durch n mögliche Tatbestände $T_1, \dots, T_n$ ersetzt. Auch diesmal müssen wir voraussetzen, daß unser Richter überzeugt ist, daß diese n Tatbestände eine erschöpfende Disjunktion aller möglichen Tatbestände und die r Maßnahmen eine erschöpfende Disjunktion aller möglichen Maßnahmen darstellen. Die daraus sich ergebende Konsequenzenmatrix kürzen wir wiederum mit $[R_{ik}]$ ab.

Das verallgemeinerte Schema hat also die folgende Gestalt:

	T_1	T_2	...	T_n
M_1	R_{11}	R_{12}	...	R_{1n}
M_2	R_{21}	R_{22}	...	R_{2n}
.	.	.		.
.	.	.		.
.	.	.		.
M_r	R_{r1}	R_{r2}		R_{rn}

mit $r \geq n^1)$

Die obige Konsequenzenmatrix schildert also die Entscheidungssituation in der Interpretation des Richters, d.h. sie repräsentiert die subjektive Überzeugung dieses Richters, der sowohl die möglichen Tatbestände wie auch die möglichen Maßnahmen, die durch den gesetzlichen Strafraum beschränkt sind, subjektiv bewertet.

Da jede Maßnahme zusammen mit einem Tatbestand die Konsequenz eindeutig festlegt, können wir jede Maßnahme als Funktion auffassen, die jedem möglichen Tatbestand T_k eine Konsequenz R_{ik} zuordnet, d.h. es gilt $M_i(T_k) = R_{ik} = (M_i, T_k)$.

Die Präferenzordnung zwischen Konsequenzen

Bevor wir zur Erläuterung der Nützlichkeitsmatrix übergehen, wollen wir annehmen, daß X in der Lage ist, anzugeben, ob er eine Konsequenz einer anderen vorzieht oder ob er irgendwelchen Konsequenzen gegenüber (mindestens zwei) indifferent ist, d.h. X muß in der Lage sein, alle möglichen Konsequenzen bzw. Resultate seiner Entscheidungen "schwach zu ordnen". Daß solche Präferenzen bzw. Wertordnungen in

1) $r \geq n$, weil auf einen Tatbestand mehrere gesetzlich festgelegte Maßnahmen (Gefängnisstrafe oder Geldstrafe, oder aber auch zusätzlich eine Nebenstrafe usw.) folgen können, aus denen dann nur eine ausgewählt wird.

der juristischen Praxis vorhanden sind, ist allgemein bekannt: Man wird in der Regel die gerechte Strafe für einen entsprechenden Tatbestand höher bewerten als dieselbe Strafe etwa für einen minder schweren Fall. Daß solche Präferenzen nicht bei allen Richtern identisch zu sein brauchen, ist auch offensichtlich: Ein Richter, der von einer Vergeltungstheorie bzw. von einer generalpräventiven Theorie ausgeht, wird anders urteilen als einer, der z.B. eine Gerechtigkeits- bzw. eine Resozialisationsauffassung vertritt.

Allgemein impliziert die schwache Ordnung folgende Annahmen: Das Entscheidungsobjekt kann bei allen möglichen Paaren von Resultaten angegeben, ob es ein Resultat vorzieht oder ob es beiden Resultaten gegenüber indifferent ist ("Vollständigkeit der Ordnung") Das Entscheidungsobjekt zeigt außerdem eine "Transitivität" innerhalb seiner Präferenzurteile: wenn es das Resultat R_1 dem Resultat R_2 , dieses wiederum dem Resultat R_3 vorzieht, so fordert die Transitivität, daß das Entscheidungsobjekt R_1 auch dem Resultat R_3 vorzieht.

Die Aussage, daß das Entscheidungsobjekt alle möglichen Konsequenzen schwach zu ordnen vermag, kann also dahingehend präzisiert werden, daß eine "Präferenz-Indifferenz-Relation $\preceq$ " (zu lesen als: "wird nicht vorgezogen")¹⁾ auf der Menge $\mathcal{R}$ der möglichen Konsequenzen gegeben ist mit folgenden Eigenschaften:

- (1) Für je zwei Konsequenzen $R_1, R_2 \in \mathcal{R}$ gilt entweder $R_1 \prec R_2$ oder $R_2 \prec R_1$ oder $R_1 \sim R_2$ (Vollständigkeit der Ordnung).
- (2) Wenn $R_1 \preceq R_2$ und $R_2 \preceq R_3$, dann $R_1 \preceq R_3$ (Transitivität der Präferenz- und Indifferenzrelation).

Die Präferenz- oder Wertordnung²⁾ des Entscheidungsobjekts bzgl. der Konsequenzen kann durch eine Nutzenfunktion abgebildet werden, d.h. jedem möglichen Resultat kann eindeutig (bis auf positiv lineare Transformationen) eine reelle Zahl, der sog. "Nutzen" zugeordnet werden. Es gilt:

$$R_1 \preceq R_2 \leftrightarrow u(R_1) \leq u(R_2)^{3)}$$

Wie eine solche Zuordnung des numerischen Nutzens konkret funktioniert, wollen wir an dieser Stelle nicht erörtern. Die Metrisierung:

1) Entsprechend können wir definieren: $R_1 \prec R_2 \leftrightarrow \neg(R_2 \preceq R_1)$, zu lesen als: " R_2 wird R_1 vorgezogen" und $R_1 \sim R_2 \leftrightarrow R_1 \preceq R_2$ und $R_2 \preceq R_1$, zu lesen als: " R_1 und R_2 sind gleichwertig".

2) auch "qualitative bzw. nichtmetrische Nutzenordnung" genannt; vgl. Stegmüller, S.298.

3) $\leq$ ist die Ordnungsrelation zwischen reellen Zahlen.

der qualitativen Nutzenordnung der Konsequenzen wird ausführlich im Anhang behandelt. Hier soll lediglich zum Ausdruck gebracht werden, daß jede beliebige reelle Zahl als "Nutzen" des Resultats fungieren kann; es gilt jedoch: Wenn das Resultat R_1 einem anderen Resultat R_2 vorgezogen wird, dann muß dem Resultat R_1 auch eine größere Zahl als dem Resultat R_2 zugeordnet werden.

Die Nützlichkeitsmatrix (oder Wünschbarkeitsmatrix)

Wir gehen davon aus, daß unser Richter X allen möglichen Konsequenzen R_{ik} "subjektive Werte" oder Nutzen zuordnet, die durch reelle Zahlen charakterisiert werden können. Setzt man für jede mögliche Konsequenz deren subjektiven Nutzen ein, so gewinnt man analog dem Beispiel aus dem 1. Kapitel eine Nützlichkeitsmatrix.

Angenommen, X geht von einer Vergeltungsstrafe mit generalpräventiver Wirkung aus, d.h. er sieht in der Vergeltung das beste Präventivmittel zur Verbrechensverhütung; er sei außerdem davon überzeugt, daß eine Strafe sowohl Abschreckungswirkung auf die Öffentlichkeit wie auch Sühnewirkung auf den Täter haben müsse. Für einen solchen Richter wird offenbar "Strafe für einen Dieb" (in dem auf S.287 angegebenen Beispiel) den höchsten Wert und "keine Strafe für einen Dieb" den niedrigsten Wert haben. Die anderen Konsequenzen "Strafe für einen Nicht-Dieb" und "keine Strafe für einen Nicht-Dieb" mögen irgendwo dazwischen liegen¹⁾. Wir nehmen an, daß für X "keine Strafe für einen Nicht-Dieb" den größeren Nutzenwert aufweist als "Strafe für einen Nicht-Dieb". Beschränken wir uns auf das Intervall $[0,1]$, dann kann die Nützlichkeitsmatrix z.B. wie folgt lauten:

1) Diese Voraussetzung verstößt insofern gegen die Rechtspraxis, als jeder Wert > 0 für "Strafe für einen Nicht-Dieb" den Grundsatz "keine Strafe ohne Verbrechen" verletzt. Es spricht jedoch viel dafür, daß eine generalpräventive Theorie mit einem derartigen Wert verträglich ist, denn mit einer solchen Theorie ist auch die Auffassung vereinbar, daß es nicht mehr auf den "wahren" Täter ankommt, sondern allein auf die Abschreckung potentieller Täter. Letzteres kann aber auch durch die Bestrafung eines Unschuldigen erreicht werden. Der obigen Forderung wird übrigens in keinem der von der juristischen Seite vorgeschlagenen Entscheidungsmodelle Rechnung getragen. Auch die empirischen Untersuchungen (v.a. die von Marshall & Wise, vgl. S.157ff dieser Arbeit) zeigen, daß viele Personen diesbezüglich in der Praxis nicht besonders zurückhaltend sind.

$$\begin{bmatrix} 1 & 0.3 \\ 0 & 0.6 \end{bmatrix}$$

Das allgemeine Schema der Nützlichkeiten ergibt sich analog wie im 1. Kapitel (§1). Wir kürzen diese Matrix wiederum mit $[u(R_{ik})]$ bzw. $[u_{ik}]$ ab. Den Wert der Funktion u für R_{ik} nennen wir auch diesmal die k te (subjektive) Nützlichkeit" oder den k ten Nutzen von R_{ik} für das Entscheidungssubjekt X ". Die Nützlichkeiten kann man, wie bereits erwähnt, der Größe nach ordnen und auf diese Weise die "numerische Ordnung" der Konsequenzen erhalten.

Die Wahrscheinlichkeitsmatrix

Es wird hier wiederum vorausgesetzt, daß X die Wahrscheinlichkeit der Realisierung jedes einzelnen der n Tatbestände zu beurteilen vermag, d.h. X kann für jeden Tatbestand T_k die subjektive Wahrscheinlichkeit seiner Realisierung $P(T_k)$, $k = 1, \dots, n$, angeben. X muß also die Wahrscheinlichkeiten für Realisierung jedes Tatbestands kennen, um zu einer Entscheidung darüber zu gelangen, welche der möglichen Maßnahmen zu wählen ist. Wie wir früher festgestellt haben, kann es sich im juristischen Fall nur um eine handlungsunabhängige Wahrscheinlichkeitsmatrix handeln¹⁾. Die Wahrscheinlichkeit der Realisierung eines beliebigen Tatbestands T_k ist nämlich davon unabhängig, welche Handlung (Maßnahme) vollzogen wird. Im juristischen Fall reicht es also aus, nur die n Werte $P(T_1) = p_1, \dots, P(T_n) = p_n$ (d.h. die Wahrscheinlichkeiten für die Tatbestände) zu kennen und diese für jede der möglichen r Maßnahmen zu übernehmen. Dagegen kann die Situation nicht eintreten, daß die Wahrscheinlichkeit dafür, daß ein Tatbestand (z.B. "Mord") erfüllt ist, auch davon abhängt, welche Maßnahme (z.B. "lebenslange Gefängnisstrafe") vollzogen wird.

Im 1. Kapitel haben wir allerdings nicht nur die handlungsabhängige, sondern auch die handlungsunabhängige Wahrscheinlichkeitsmatrix aus Gründen der Übersichtlichkeit in der Weise gebildet, daß wir die

1) Im 1. Kap. haben wir zwischen einer handlungsabhängigen und einer handlungsunabhängigen Wahrscheinlichkeitsmatrix unterschieden. In der Konzeption von Jeffrey ist eine solche Unterscheidung selbstverständlich. Die Konzeption von Savage, die unserer Theorie zugrunde liegt, kennt jedoch nur handlungsunabhängige Ereignisse, denen man Wahrscheinlichkeiten zuordnet. Diese Einschränkung wird von der Rechtspraxis aber trivialerweise erfüllt. Zum Savage-Modell vgl. S. 271ff dieser Arbeit.

Wahrscheinlichkeitsverteilung für n Umstände m -mal (d.h. für jede der m möglichen Handlungen) untereinander geschrieben haben. Dies ist aus zwei Gründen mißverständlich. Erstens könnte man meinen, daß die Wahrscheinlichkeiten den Konsequenzen zukommen¹⁾ und zweitens ist eine solche Schreibweise mathematisch unzulässig²⁾. Um derartige Mißdeutungen zu vermeiden, wollen wir uns an die in der Mathematik übliche Schreibweise halten und die Wahrscheinlichkeitsverteilung für n Tatbestände als eine einzeilige Matrix, einen sog. Vektor, darstellen. Die Wahrscheinlichkeitsmatrix bzw. der Wahrscheinlichkeitsvektor hat also folgendes Aussehen:

$$[p_1, p_2, \dots, p_n]$$

Ist p in unserem Beispiel die Wahrscheinlichkeit, daß der Angeklagte den Diebstahl begangen hat, so ist die Wahrscheinlichkeit dafür, daß er den Diebstahl nicht begangen hat, $1 - p$. Mit $p = 0.7$ hat der Wahrscheinlichkeitsvektor die konkrete Gestalt:

$$[0.7, 0.3]$$

Die Präferenzordnung zwischen Maßnahmen

Mit der Nützlichkeitsmatrix und dem Wahrscheinlichkeitsvektor kann X nun die Präferenzordnung der Maßnahmen ermitteln. Dies geschieht, indem X den erwarteten Nutzen für jede Maßnahme berechnet. Die Operation, die einer solchen Berechnung zugrunde liegt, heißt Matrizenmultiplikation.

1) Wir haben im 1.Kap. §1, die Wahrscheinlichkeiten für die Umstände automatisch (Zeile für Zeile) auf die Konsequenzen übertragen, was damals keine Schwierigkeit darstellte: Das Entscheidungsmodell war noch zu allgemein (es kannte noch keine Savage-Restriktion hinsichtlich der Zuordnung von Wahrscheinlichkeiten), man konnte ohne weiteres auch über die Wahrscheinlichkeiten der Konsequenzen reden (was in der Jeffrey-Theorie auch der Fall ist). Jetzt hingegen müssen wir vorsichtig sein und die obige Wahrscheinlichkeitsverteilung auch optisch als eine solche für die möglichen Tatbestände zu erkennen geben.

2) Es macht nämlich (bei der Matrizenmultiplikation) einen Unterschied, ob man eine Matrix mit einer $m \times n$ - oder mit einer $1 \times n$ -Matrix (genauer: mit einem $n \times 1$ -Spaltenvektor) multipliziert. Im letzteren Fall ist das Ergebnis auch ein Vektor, im ersten hingegen nicht (falls $m > 1$). Daß wir damals trotzdem ein richtiges Ergebnis erhielten, auch mit der mathematisch unkorrekt angeschriebenen Wahrscheinlichkeitsmatrix, verdanken wir dem Umstand, daß wir die Operation der Matrizenmultiplikation richtig (d.h. in bezug auf einen Wahrscheinlichkeitsvektor) angewandt haben.

Angenommen, die Nützlichkeitsmatrix sehe so aus:

$$\begin{bmatrix} u_{11} & u_{12} & \dots & u_{1n} \\ u_{21} & \dots & \dots & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ u_{r1} & \dots & \dots & u_{rn} \end{bmatrix} = [u_{ik}], \quad \begin{matrix} i = 1, \dots, r \\ k = 1, \dots, n \end{matrix}$$

Ferner sei der Wahrscheinlichkeitsvektor

$$[p_1, p_2, \dots, p_n]$$

gegeben.

Die Matrizenmultiplikation geschieht in der Weise, daß man zuerst links die Nützlichkeitsmatrix und rechts (als Spaltenvektor) die Wahrscheinlichkeitsmatrix anschreibt; anschließend multipliziert man jede Nützlichkeit in der ersten Zeile der Nützlichkeitsmatrix mit der entsprechenden Wahrscheinlichkeit in der Wahrscheinlichkeitsmatrix und addiert die so erhaltenen Werte zueinander. Das Ergebnis schreibt man als die erste Zeile in der neuen Matrix, der sog. Ergebnismatrix an. Diese Operation wiederholt man für jede Zeile. Als Ergebnis erhalten wir wiederum eine Matrix, im obigen Fall eine $r \times 1$ -Matrix (also einen Vektor).

Die Matrizenmultiplikation läßt sich formal so schreiben:

$$\begin{bmatrix} u_{11} & u_{12} & \dots & u_{1n} \\ u_{21} & \dots & \dots & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ u_{r1} & \dots & \dots & u_{rn} \end{bmatrix} \cdot \begin{bmatrix} p_1 \\ p_2 \\ \cdot \\ \cdot \\ \cdot \\ p_n \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^n u_{1k} p_k \\ \sum_{k=1}^n u_{2k} p_k \\ \cdot \\ \cdot \\ \cdot \\ \sum_{k=1}^n u_{rk} p_k \end{bmatrix} =$$

$$= [\sum_{k=1}^n u_{ik} p_k]_{i=1, \dots, r} = [E_u(M_i)]_{i=1, \dots, r}.$$

Jede Zeile der obigen Matrix repräsentiert den Erwartungswert des Nutzens (bzw. den subjektiven Wert) jeder einzelnen der r Maßnahmen. Der Erwartungswert (bzw. der subjektive Wert) $E_u(M_i)$ der Maßnahme M_i ist also

$$E_u(M_i) = \sum_{k=1}^n u_{ik} p_k$$

Aus der Matrix gewinnt man r subjektive Werte $E_u(M_1), \dots, E_u(M_r)$. Die Größenordnung zwischen diesen Werten heißt, wie früher (1. Kap. §1), numerische Präferenzordnung der Maßnahmen.

In unserem Beispiel erhalten wir somit:

$$\begin{bmatrix} 1 & 0.3 \\ 0 & 0.6 \end{bmatrix} \cdot \begin{bmatrix} 0.7 \\ 0.3 \end{bmatrix} = \begin{bmatrix} 1 \cdot 0.7 + 0.3 \cdot 0.3 = 0.79 \\ 0 \cdot 0.7 + 0.6 \cdot 0.3 = 0.18 \end{bmatrix}$$

Die subjektiven Werte lauten also:

$$E_u(M_1) = 0.79 \quad \text{und} \quad E_u(M_2) = 0.18.$$

Die Regel von Bayes

Wir gehen nun von der in der rationalen Entscheidungstheorie gängigen Annahme aus, daß ein rationales Entscheidungssubjekt "seine Entscheidung in der Weise trifft, daß es etwas maximiert"¹⁾ Im Fall von Entscheidungen unter Sicherheit nimmt man an, daß das Entscheidungssubjekt einen bestimmten Nutzen (seinen eigenen oder den der Gesellschaft) maximiert. Aus der Menge der möglichen Handlungen wählt man diejenige aus, die in der subjektiven Präferenzordnung allen anderen vorgezogen wird, d.h. deren Nutzen am höchsten ist. Im Fall von Entscheidungen unter Risiko (mit denen wir es hier zu tun haben) läßt sich der Nutzen offensichtlich nicht so ohne weiteres maximieren, denn man weiß ja nicht, ob man diesen Nutzen überhaupt erhält. Statt dessen wird hier der Erwartungswert des aus der Handlung resultierenden Nutzens (genauer: des Nutzens, der sich aus allen möglichen Konsequenzen ergibt mit den Wahrscheinlichkeiten der entsprechenden Umstände als Wichtungskoeffizienten) maximiert. Aus der Menge der möglichen Handlungen wählt man dann diejenige aus, die in der subjektiven Präferenzordnung den höchsten Erwartungswert des Nutzens hat²⁾

Wir schreiben hier noch einmal (diesmal jedoch formal) die Regel von Bayes an.

Bezeichnet man mit M_i die i-te Maßnahme, so lautet die Bayes-Regel für die Maximierung des Erwartungswertes des juristischen Nutzens:

$$M_i \succ M_j \leftrightarrow \sum_{k=1}^n u_{ik} p_k \geq \sum_{k=1}^n u_{jk} p_k \leftrightarrow E_u(M_i) \geq E_u(M_j) \quad \text{für alle } i.$$

Diese Regel besagt: Wähle unter den möglichen Maßnahmen diejenige aus, welche in der numerischen Präferenzordnung den höchsten subjektiven Wert besitzt.

(Im obigen Beispiel muß X also die erste Maßnahme wählen).

1) Vgl. Kirsch, Entscheidungsprozesse I, S.40.

2) Für die Entscheidungssituation unter Unsicherheit, in welcher keine Wahrscheinlichkeiten angebar sind, sind inzwischen mehrere Entscheidungsregeln formuliert worden (die jeweils zu unterschiedlichen Entscheidungspräferenzen führen). Auf diese wird hier nicht näher eingegangen.

Unsere Konzeption können wir nun auch abstrakt darstellen. Diese Möglichkeit sei hier kurz erläutert.

Wir fassen Maßnahmen, die einem Entscheidenden X offenstehen, in eine Menge $\mathcal{M}$ zusammen. Juristische Umstände nannten wir früher in Anlehnung an den juristischen Sprachgebrauch Tatbestände. Im Grunde genommen handelt es sich hier jedoch um sog. "elementare Tatbestände" ("elementare Sachverhalte"). Die elementaren Sachverhalte bilden wiederum eine Menge, die wir als Sachverhaltsmenge S (bzw. als Raum der möglichen Sachverhalte) bezeichnen. Eine Menge von elementaren Sachverhalten nennt man "Ereignis". Ein Ereignis kann also als jene Teilmenge der Sachverhaltsmenge dargestellt werden, die genau diejenigen Sachverhalte enthält, in denen dieses Ereignis gilt. Ereignisse sind Elemente von $\text{Pot}(S)$, die abgeschlossen ist¹⁾, d.h. es gilt:

- (1) $A \in \text{Pot}(S) \rightarrow \bar{A} \in \text{Pot}(S)$;
- (2) $A, B \in \text{Pot}(S) \rightarrow A \cap B \in \text{Pot}(S)$;
- (3) $A, B \in \text{Pot}(S) \rightarrow A \cup B \in \text{Pot}(S)$.

($\text{Pot}(S)$ ist somit eine boolesche Algebra).

Wir können die juristische Entscheidungssituation (JES) auf folgende Weise charakterisieren:

Ein 8-Tupel $\langle X, S, \mathcal{T}, \mathcal{R}, \mathcal{M}, P, u, E_u \rangle$ kennzeichnet eine JES genau dann, wenn gilt:

- (1) X ist ein Entscheidungssubjekt oder Individuum (der Richter).
- (2) S ist eine nicht leere endliche Menge der möglichen elementaren Sachverhalte (elementaren Tatbestände).
- (3) $\mathcal{T}$ ist eine nicht leere endliche Menge disjunkter, erschöpfender und maßnahmenunabhängiger Ereignisse; diese heißen Tatbestände²⁾ und sind ihrerseits Mengen von elementaren Sachverhalten aus S .
- (4) $\mathcal{R}$ ist eine nicht leere endliche Menge der möglichen Konsequenzen.
- (5) $\mathcal{M}$ ist eine nicht leere Menge von Funktionen von S in $\mathcal{R}$; diese Funktionen repräsentieren die dem Richter offenstehenden Maßnahmen³⁾.
- (6) P ist eine subjektive Wahrscheinlichkeitsfunktion auf $\text{Pot}(S)$; es gilt für jedes $T \subseteq S$: $P(T) = \sum_{A \in \mathcal{T}} P(A)$.
- (7) u ist eine subjektive Bewertungsfunktion (Nutzenfunktion) von $\mathcal{R}$ in $\mathbb{R}$.
- (8) E_u ist eine auf $\mathcal{M}$ definierte Funktion, für die gilt:
für alle $M \in \mathcal{M}$ $E_u(M) = \sum_{A \in S} u(M(A)) \cdot P(A)$.

Die mit diesem Modell verbundene normative Regel besagt:

Wird X durch das Entscheidungsmodell $\langle X, S, \mathcal{T}, \mathcal{R}, \mathcal{M}, P, u, E_u \rangle$ charakterisiert, so soll X ein $M \in \mathcal{M}$ wählen, für das $E_u(M)$ maximal ist.

1) Vgl. S.13 dieser Arbeit.

2) Auf die Anführung von $\mathcal{T}$ kann verzichtet werden. In Hinblick auf die juristische Unterscheidung zwischen "Grundtatbeständen" (in unserer jetzigen Terminologie: "elementare Sachverhalte") und "qualifizierten Tatbeständen" (hier: "Ereignisse") scheint die Beibehaltung von $\mathcal{T}$ allerdings ratsam. Auf diese Weise kann ein Jurist leicht erfahren, welchen Elementen im Modell seine Tatbestände formal entsprechen. Auf diesen Punkt werden wir im nächsten Abschnitt näher eingehen.

3) Zur Darstellung von Handlungen als Funktionen vgl. S.271, Fn.3.

1.2. Qualifizierung des Tatbestandes

In vielen juristischen Fällen wird nach der Feststellung des sog. Grundtatbestandes überprüft, ob eine Qualifizierung vorliegt oder nicht. So ist z.B. der Grundtatbestand von "schwerer Diebstahl", "Diebstahl mit Waffen" etc. der Tatbestand "Diebstahl"; von "Tötung auf Verlangen", "fahrlässige Tötung", "Mord" der Tatbestand "Tötung" ("Totschlag"). Welche Probleme entstehen für uns, wenn wir nicht nur die Grundtatbestände, sondern auch qualifizierte Tatbestände einer Entscheidungssituation zugrunde legen?

Angenommen, X muß sich entscheiden, ob er den Angeklagten für "Diebstahl" oder "Diebstahl mit Waffen" verurteilen oder ihn freisprechen soll. Es kommen also drei Tatbestände in Frage: "Diebstahl", "Diebstahl mit Waffen" und "kein Diebstahl". Nun könnte man meinen, daß "Diebstahl mit Waffen" nichts anderes ist als ein zusammengesetzter Sachverhalt, dessen Wahrscheinlichkeit daher kleiner sein müßte als die für "Diebstahl". Die Wahrscheinlichkeit für die drei Tatbestände würde sich dann nicht zu 1 addieren, bzw. größer als 1 sein, was ein Zeichen dafür wäre, das etwas nicht stimmt. Wir müssen somit sehen, wie wir dieses Mißverständnis auflösen.

Als erstes wird hier gezeigt, wie man aus der Sachverhaltsmenge, die aus dem Grundtatbestand "Diebstahl" und seiner Negation "kein Diebstahl" besteht, rein formal die obige Tatbestandsmenge erhält.

Sei die Sachverhaltsmenge {Diebstahl, kein Diebstahl} gegeben. Diese Menge ist zu der folgenden Menge logisch äquivalent: {Diebstahl&Waffen, Diebstahl&keine Waffen, kein Diebstahl&Waffen, kein Diebstahl&keine Waffen}¹⁾

Aus dieser Menge erhalten wir die gesuchte Tatbestandsmenge, indem wir ihre Teilmengen mit den juristischen Tatbeständen identifizieren. So können wir "Diebstahl&Waffen" mit dem juristischen Tatbestand "Diebstahl mit Waffen", "Diebstahl&keine Waffen" mit "Diebstahl" und ("kein Diebstahl&Waffen" oder "kein Diebstahl&keine

1) Allgemein gilt: Sind n verschiedenen Merkmale gegeben, so gibt es 2^n verschiedene Kombinationen dieser Merkmale (diese Kombinationen identifizieren wir mit möglichen Ereignissen).

Waffen") mit "kein Diebstahl"¹⁾ identifizieren.

So kann man also ohne Schwierigkeiten aus einer ursprünglichen Sachverhaltsmenge, welche elementare Tatbestände enthält (bzw. aus der ursprünglichen Entscheidungssituation) eine Tatbestandsmenge mit neuen qualifizierten Tatbeständen (bzw. eine neue Entscheidungssituation, in der auch die Qualifizierung der Tatbestände berücksichtigt wird) herleiten, indem man eine der ersten Menge logisch äquivalente Menge konstruiert, die unter ihren Elementen Merkmale enthält, welche die neuen Tatbestände mitkonstituieren, und anschließend die Teilmengen dieser Menge mit diesen neuen Tatbeständen identifiziert²⁾

Nehmen wir nun an, die Wahrscheinlichkeit für den Grundtatbestand "Diebstahl" beträgt 0.9. Die Wahrscheinlichkeit für den negativen Tatbestand "kein Diebstahl" ist dann 0.1.

Um die Verwechslung zwischen einem Grundtatbestand (elementaren Sachverhalt) "Diebstahl" und einem abgeleiteten Tatbestand "Diebstahl" ($\hat{=}$ "Diebstahl & keine Waffen") zu vermeiden, wollen wir nur den ersteren als Grundtatbestand bezeichnen; ähnliches gilt auch für alle anderen Grund- und abgeleiteten Tatbestände).

Zur Berechnung der Wahrscheinlichkeit für die qualifizierten Tatbestände müssen wir außerdem die Wahrscheinlichkeit für das Merkmal "Waffen" kennen (Nehmen wir an, daß die beiden Ereignisse "Diebstahl" und "Waffen" voneinander unabhängig sind). Sei eine solche ebenfalls 0.9, dann ist die Wahrscheinlichkeit für "keine Waffen" 0.1. Mit diesen Werten erhalten wir für die drei Tatbestände: "Diebstahl", "kein Diebstahl" und "Diebstahl mit Waffen" folgende Wahrscheinlichkeiten:

$$P(\text{Diebstahl \& Waffen}) = 0.9 \cdot 0.9 = 0.81;$$

$$P(\text{Diebstahl}) = P(\text{Diebstahl \& keine Waffen}) = 0.9 \cdot 0.1 = 0.09;$$

$$P(\text{kein Diebstahl}) = P((\text{kein Diebstahl \& Waffen}) \vee (\text{kein Diebstahl \& keine Waffen})) = 0.1 \cdot 0.9 + 0.1 \cdot 0.1 = 0.1.$$

Der Wahrscheinlichkeitsvektor lautet somit:

1) Käme zu den obigen drei Tatbeständen ein vierter, z.B. "Waffenbesitz" hinzu, so würden wir die aus vier Tatbeständen bestehende Menge gewinnen, indem wir "kein Diebstahl & Waffen" mit "Waffenbesitz" identifizieren.

2) Jetzt zeigt sich der Nutzen der Extraanführung der Tatbestandsmenge $\supset$ und der der Unterscheidung zwischen der Sachverhaltsmenge und der Tatbestandsmenge auf S.296. Oben wurde gezeigt, daß die Elemente der Tatbestandsmenge (qualifizierte Tatbestände) als Teilmengen der modifizierten (jedoch zu der ursprünglichen logisch äqui-

[0.81, 0.09, 0.1]

Wie wir sehen, sind die juristischen Ereignisse "Diebstahl" und "kein Diebstahl" im obigen Beispiel nicht erschöpfend. Es ist also zu beachten, daß juristische Ereignisse nicht mit den gleichnamigen probabilistischen Ereignissen gleichgesetzt werden dürfen.

Man könnte einwenden, die obige Konstruktion, d.h. die Identifikation des juristischen Tatbestands "Diebstahl" mit dem Ereignis "Diebstahl & keine Waffen" und "kein Diebstahl" mit etwas noch komplizierterem sei zu künstlich und habe mit einer juristischen Beurteilung nichts zu tun; es genüge für juristische Zwecke, wenn man den qualifizierten Tatbestand "Diebstahl mit Waffen" einerseits gegen den Grundtatbestand "Diebstahl" und andererseits gegen den Sachverhalt "kein Diebstahl" abwäge. Dem ist zu entgegnen:

1. Die Wahrscheinlichkeitswerte für diese Sachverhalte würden dann lauten: 0.82, 0.9 und 0.1, so daß ihre Summe 1.82 ergäbe, im Widerspruch zu den Wahrscheinlichkeitstheoretischen Axiomen.
2. Die Wahrscheinlichkeit für "Diebstahl mit Waffen" (oder aber auch für einen beliebigen qualifizierten Tatbestand) ist dann immer kleiner als die Wahrscheinlichkeit für den Grundtatbestand: somit dürfte man konsequenterweise niemals auf Grund des qualifizierten Tatbestandes verurteilen;
3. Die obige Vorgehensweise würde gegen gewisse Rationalitätsforderungen verstoßen, etwa gegen die, daß man alle logisch möglichen Ereignisse beurteilen soll. Im obigen Beispiel waren es zwei Merkmale ("Diebstahl" und "Waffen"); es sind somit $2^2 = 4$ Ereignisse logisch möglich (die aus den Konjunktionen von bestehenden oder nicht bestehenden Merkmalen gebildet werden). Unser Trick bestand darin, daß wir juristische Tatbestände mit diesen Ereignissen bzw. mit Adjunktionen von diesen Ereignissen identifizierten, zum einen, um den Begriff "qualifizierter Tatbestand" formal zu präzisieren, zum anderen, um überhaupt in unserem Modell von solchen Objekten sinnvoll reden zu können.

Wir rekonstruieren nun die Entscheidungssituation, der die obigen drei Tatbestände zugrunde liegen.

valenten) Sachverhaltsmenge bzw. als Elemente der Potenzmenge über dieser (neuen) Menge aufgefaßt werden können. Somit erfährt der juristische Begriff "qualifizierter Tatbestand" in unserem Modell eine formale Klärung.

Den Tatbeständen "Diebstahl mit Waffen" (T_1), "Diebstahl" (T_2) und "kein Diebstahl" (T_3) entsprechen laut StGB folgende Maßnahmen: "nach §244 StGB" (M_1), "nach §242 StGB" (M_2), "Freispruch" (M_3). Die Konsequenzenmatrix sieht also so aus:

	T_1	T_2	T_3
M_1 : §244	(M_1, T_1)	(M_1, T_2)	(M_1, T_3)
M_2 : §242	(M_2, T_1)	(M_2, T_2)	(M_2, T_3)
M_3 : Freispruch	(M_3, T_1)	(M_3, T_2)	(M_3, T_3)

Die Nützlichkeitsmatrix möge die Gestalt haben:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0.5 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

D.h. die gerechten Behandlungen (d.h. tatschuldangemessene Bestrafung¹⁾ für den entsprechenden Tatbestand und Freispruch für einen Nicht-Täter) erhalten die höchsten Nützlichkeiten; die mildere als dem Tatbestand entsprechende Bestrafung ("nach §242 (wegen Diebstahls) für Diebstahl mit Waffen") erhält die positive Nützlichkeit 0.5; "die Bestrafung des Nicht-Täters" erhält in jedem Fall den niedrigsten Wert 0 und außerdem wird "Freispruch für den Täter" in beiden Fällen auch mit 0 bewertet.

Die Multiplikation der obigen Nützlichkeitsmatrix mit dem Wahrscheinlichkeitsvektor $[0.81, 0.09, 0.1]$ ergibt folgende erwartete Nutzen:

$$E_u(M_1) = 0.81, E_u(M_2) = 0.495 \text{ und } E_u(M_3) = 0.1.$$

Also muß hier die Strafe nach §244 StGB (wegen "Diebstahls mit Waffen") gewählt werden.

Nun sei angenommen, daß zu den obigen drei Tatbeständen - weil eine zusätzliche Qualifikation (wie etwa: "Drohung mit gegenwärtiger Gefahr für Leib oder Leben", wir kürzen das mit "Bedrohung" ab) vorliegt - noch die Tatbestände "Raub" und "schwerer Raub" hinzukommen. Es sind somit 5 Tatbestände, die wir überprüfen wollen: "schwerer Raub" (T_1), "Raub" (T_2), "Diebstahl mit Waffen" (T_3), "Diebstahl" (T_4) und "kein Diebstahl" (T_5). Wir wollen die Tatbestandsmenge

1) Unter einer "tatschuldangemessenen Strafe" versteht man die gesetzlich vorgeschriebene Bestrafung für den entsprechenden Tatbestand; so ist zwar die Strafe nach §249 StGB (wegen Raubes) für den Tatbestand "Raub" tatschuldangemessen, nicht dagegen für "schweren Raub", "Diebstahl", "Mord" usw. Ob die tatschuldangemessene Strafe im ethischen Sinne gerecht ist, sei hier dahingestellt. Hier wird "Gerechtigkeit" nur im formalen Sinne verstanden.

$\{T_1, T_2, T_3, T_4, T_5\}$ auf die gleiche Weise wie vorher rekonstruieren. Zunächst ermitteln wir, wieviele unterschiedliche Merkmale in Frage kommen. Im obigen Beispiel sind es drei: "Diebstahl", "Waffen" und "Bedrohung". Daraus ergibt sich, daß $2^3 = 8$ verschiedene Ergebnisse möglich sind:

{Diebstahl&Waffen&Bedrohung, Diebstahl&Waffen&keine Bedrohung, Diebstahl&keine Waffen&Bedrohung, Diebstahl&keine Waffen&keine Bedrohung, kein Diebstahl&Waffen&Bedrohung, kein Diebstahl&Waffen&keine Bedrohung, kein Diebstahl&keine Waffen&Bedrohung, kein Diebstahl&keine Waffen&keine Bedrohung}.

Diese Menge ist übrigens nicht nur der Sachverhaltsmenge {Diebstahl, kein Diebstahl}, sondern auch der Tatbestandsmenge {Diebstahl&Waffen, Diebstahl&keine Waffen, kein Diebstahl&Waffen, kein Diebstahl&keine Waffen} logisch äquivalent.

Wir können das leicht nachprüfen: Seien A, B und C Abkürzungen für "Diebstahl", "Waffen" und "Bedrohung". Es gilt:

$$\begin{aligned} & ((A \wedge B \wedge C) \vee (A \wedge B \wedge \neg C)) \vee ((A \wedge \neg B \wedge C) \vee (A \wedge \neg B \wedge \neg C)) \vee ((\neg A \wedge B \wedge C) \vee (\neg A \wedge B \wedge \neg C)) \\ & \vee ((\neg A \wedge \neg B \wedge C) \vee (\neg A \wedge \neg B \wedge \neg C)) \leftrightarrow (A \wedge B) \vee (A \wedge \neg B) \vee (\neg A \wedge B) \vee (\neg A \wedge \neg B) \\ & \leftrightarrow (A) \vee (\neg A). \end{aligned}$$

(den letzten Ausdruck erhalten wir durch sukzessives Zusammenziehen der Glieder, z.B. ist $((A \wedge B \wedge C) \vee (A \wedge B \wedge \neg C))$ logisch äquivalent zu $(A \wedge B)$).

Wiederum identifizieren wir juristische Tatbestände mit den Teilmengen der angeführten Menge:

$T_1 \hat{=}$ "Diebstahl&Waffen&Bedrohung";

$T_2 \hat{=}$ "Diebstahl&keine Waffen&Bedrohung";

$T_3 \hat{=}$ "Diebstahl&Waffen&keine Bedrohung";

$T_4 \hat{=}$ "Diebstahl&keine Waffen&keine Bedrohung";

$T_5 \hat{=}$ "kein Diebstahl&Waffen&Bedrohung" oder "kein Diebstahl&Waffen&keine Bedrohung" oder "kein Diebstahl&keine Waffen&Bedrohung" oder "kein Diebstahl&keine Waffen&keine Bedrohung" =
= "kein Diebstahl".

Wären wir außerdem an der Beurteilung der Tatbestände "Waffenbesitz" oder "Bedrohung" interessiert, so würden wir noch zwei zusätzliche Tatbestände "Waffenbesitz" (T_6) und "Bedrohung" (T_7) in Betracht ziehen, wobei T_5 nun mit "kein Diebstahl&keine Waffen&keine Bedrohung", T_6 mit "kein Diebstahl&Waffen&Bedrohung" oder "kein Diebstahl&Waffen&keine Bedrohung" und T_7 mit "kein Diebstahl&keine Waffen&Bedrohung" identisch sein müßten. Die Gleichsetzung von T_6 und nicht T_7 mit "kein Diebstahl&Waffen&Bedrohung" erfolgt aus einem juristischen Grund: Nach §52 (Tateinheit) StGB muß die Strafe nach dem Gesetz bestimmt werden, das die schwerste Strafe androht, falls gegen mehrere Strafgesetze (in diesem Fall offensichtlich zwei: §241 StGB (Bedrohung) und §117 (WaffG)) verstoßen wurde. Da "Bedrohung" mit einer Freiheitsstrafe bis zu einem Jahr oder mit einer Geldstrafe

bestraft wird, der (unerlaubte) Waffenbesitz dagegen in manchen Fällen mit einer Freiheitsstrafe bis zu 10 Jahren, dürfte die obige Identifikation plausibel sein.

Die Konsequenzenmatrix hat also folgende Gestalt:

	T_1	T_2	T_3	T_4	T_5
M_1 : §250	(M_1, T_1)	(M_1, T_2)	(M_1, T_3)	(M_1, T_4)	(M_1, T_5)
M_2 : §249	(M_2, T_1)	(M_2, T_2)	(M_2, T_3)	(M_2, T_4)	(M_2, T_5)
M_3 : §244	(M_3, T_1)	(M_3, T_2)	(M_3, T_3)	(M_3, T_4)	(M_3, T_5)
M_4 : §242	(M_4, T_1)	(M_4, T_2)	(M_4, T_3)	(M_4, T_4)	(M_4, T_5)
M_5 : Freispruch	(M_5, T_1)	(M_5, T_2)	(M_5, T_3)	(M_5, T_4)	(M_5, T_5)

Seien nun die Wahrscheinlichkeiten für "Diebstahl" und "Waffen" wie früher, d.h. jeweils 0.9, und die Wahrscheinlichkeit für "Bedrohung" 0.5. Wir nehmen einfachheitshalber wieder die Unabhängigkeit der drei Merkmale an.

Wenn diese nicht erfüllt ist, müssen wir die Wahrscheinlichkeit für das gemeinsame Vorliegen der Merkmale mit der Formel $P(A \wedge B \wedge C) = P(A/B \wedge C) \cdot P(B/C) \cdot P(C)$ bestimmen.

Sei "A" die Abkürzung für "Diebstahl", "B" für "Bedrohung", "C" für "Waffen". dann müssen wir laut dieser Formel die Wahrscheinlichkeit für "Waffen" mit der bedingten Wahrscheinlichkeit für "Bedrohung" aufgrund "Waffen" und mit der Wahrscheinlichkeit für "Diebstahl" aufgrund "Waffen und Bedrohung" multiplizieren.

Wir erhalten also den folgenden Wahrscheinlichkeitsvektor:

$$[0.405, 0.045, 0.405, 0.045, 0.1].$$

Die Nützlichkeitsmatrix habe die Gestalt:

1	0	0	0	0
0.5	1	0	0	0
0	0	1	0	0
0	0	0	1	0
0	0	0	0	1

D.h. die gerechten Beurteilungen (tatschuldangemessene Bestrafung für den entsprechenden Tatbestand und Freispruch für den Nicht-Täter) erhalten den höchsten Nützlichkeitswert 1. Die nichttatschuldangemessenen Bestrafungen bzw. Freispruch für den Täter erhalten in jedem Fall den niedrigsten Wert 0; Ausnahme: die mildere "Bestrafung nach §249 (wegen Raubs) StGB für schweren Raub" erhält in der obigen Nützlichkeitsmatrix die Nützlichkeit 0.5.

Multiplikation der obigen Nützlichkeitsmatrix mit dem Wahrscheinlichkeitsvektor ergibt die folgenden Erwartungswerte des Nutzens:

$$E_u(M_1) = 0.405, E_u(M_2) = 0.2475, E_u(M_3) = 0.405, E_u(M_4) = 0.045$$

und $E_u(M_5) = 0.1$.

Die beiden Maßnahmen M_1 ("Bestrafung nach §250 StGB (schwerer Raub)") und M_3 ("Bestrafung nach §244 StGB (Diebstahl mit Waffen)") erhalten den gleichen höchsten Erwartungswert des Nutzens.

Ein weiteres Beispiel:

Wir nehmen an, daß der Entscheidungssituation folgende Tatbestände zugrunde liegen: "Mord" (T_1), "Totschlag" (T_2), "minder schwerer Fall von Totschlag" (T_3) und "keine Tötung" (T_4).

Für "Mord" spricht, daß der Angeklagte im Fall des Todes des Opfers ein großes Erbe erhalten hätte, so daß es wahrscheinlich ist, daß der Angeklagte das Opfer aus Habgier hätte töten können. Für "minder schweren Fall von Totschlag" spricht aber, daß das Opfer die Tat hätte auslösen können, da es den Angeklagten vermutlich mißhandelte (es sei bekannt, daß das Opfer dazu neigte). Es gibt also drei Merkmale "Tötung", "Habgier" und "Mißhandlung" (wir kürzen sie jeweils mit " E_1 ", " E_2 " und " E_3 " ab) und 8 Ereignisse:

$\{E_1 \wedge E_2 \wedge E_3, E_1 \wedge E_2 \wedge \neg E_3, E_1 \wedge \neg E_2 \wedge E_3, E_1 \wedge \neg E_2 \wedge \neg E_3, \neg E_1 \wedge E_2 \wedge E_3, \neg E_1 \wedge E_2 \wedge \neg E_3, \neg E_1 \wedge \neg E_2 \wedge E_3, \neg E_1 \wedge \neg E_2 \wedge \neg E_3\}$.

Wir identifizieren wiederum juristische Tatbestände mit den Teilmengen der angeführten Menge.

$T_1 \triangleq$ "Tötung&Habgier&keine Mißhandlung" $\triangleq E_1 \wedge E_2 \wedge \neg E_3$;

$T_2 \triangleq$ "Tötung&Habgier&Mißhandlung" oder "Tötung&keine Habgier&keine Mißhandlung" $\triangleq (E_1 \wedge E_2 \wedge E_3) \vee (E_1 \wedge \neg E_2 \wedge \neg E_3)$;

$T_3 \triangleq$ "Tötung&keine Habgier&Mißhandlung" $\triangleq E_1 \wedge \neg E_2 \wedge E_3$;

$T_4 \triangleq$ "keine Tötung&Habgier&Mißhandlung" oder "keine Tötung&Habgier&keine Mißhandlung" oder "keine Tötung&keine Habgier&Mißhandlung" oder "keine Tötung&keine Habgier&keine Mißhandlung" $\triangleq (\neg E_1 \wedge E_2 \wedge E_3) \vee (\neg E_1 \wedge E_2 \wedge \neg E_3) \vee (\neg E_1 \wedge \neg E_2 \wedge E_3) \vee (\neg E_1 \wedge \neg E_2 \wedge \neg E_3) = \neg E_1$.

Nun bestimmen wir die Wahrscheinlichkeit für die Tatbestände T_1 bis T_4 . Zwar ist anzunehmen, daß die beiden Ereignisse "Habgier" und "Mißhandlung" voneinander unabhängig sind, in bezug auf die drei Ereignisse "Tötung", "Habgier" und "Mißhandlung" ist diese Annahme jedoch unrealistisch. Wir haben also die bedingten Wahrscheinlichkeiten zu bestimmen.

Folgende Wahrscheinlichkeitswerte seien gegeben:

$$P(E_2) = P(E_3) = 0.5; P(E_1/E_2 \wedge E_3) = 0.9, P(E_1/\neg E_2 \wedge E_3) = 0.8; \\ P(E_1/E_2 \wedge \neg E_3) = 0.8 \text{ und } P(E_1/\neg E_2 \wedge \neg E_3) = 0.6^1)$$

Für die $P(T_i)$, $i = 1, \dots, 4$ ergibt sich dann:

$$P(T_1) = P(E_1 \wedge E_2 \wedge \neg E_3) = P(E_1/E_2 \wedge \neg E_3) \cdot P(E_2/\neg E_3) \cdot P(\neg E_3) = \\ = P(E_1/E_2 \wedge \neg E_3) \cdot P(E_2) \cdot P(\neg E_3)^2 = 0.2;$$

$$P(T_2) = P((E_1 \wedge E_2 \wedge E_3) \vee (E_1 \wedge \neg E_2 \wedge \neg E_3)) = P(E_1/E_2 \wedge E_3) \cdot P(E_2) \cdot P(E_3) + \\ + P(E_1/\neg E_2 \wedge \neg E_3) \cdot P(\neg E_2) \cdot P(\neg E_3) = 0.375;$$

$$P(T_3) = P(E_1 \wedge \neg E_2 \wedge E_3) = P(E_1/\neg E_2 \wedge E_3) \cdot P(\neg E_2/E_3) \cdot P(E_3) = \\ P(E_1/\neg E_2 \wedge E_3) \cdot P(\neg E_2) \cdot P(E_3) = 0.2;$$

$$P(T_4) = P(\neg E_1) = 1 - [P(T_1) + P(T_2) + P(T_3)] = 0.225.$$

Der Wahrscheinlichkeitsvektor lautet somit:

$$[0.2, 0.375, 0.2, 0.225].$$

4 verschiedene Maßnahmen kommen in Frage:

M_1 : "nach §211 StGB (wegen Mords)"; M_2 : "nach §212 StGB (wegen Totschlags)"; M_3 : "nach §213 StGB (wegen eines minder schweren Falls von Totschlag)" und M_4 : "Freispruch".

Wir schreiben die Konsequenzenmatrix in Kurzform an:

$$\begin{array}{c} T_1 \quad T_2 \quad T_3 \quad T_4 \\ M_1 \\ M_2 \quad [M_i, T_j], \quad i=1, \dots, 4 \\ M_3 \quad \quad \quad j=1, \dots, 4 \\ M_4 \end{array}$$

Die Nützlichkeitsmatrix möge so aussehen:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0.8 \end{bmatrix}$$

D.h. die tatschuldangemessene Bestrafung für den Mord wird am höchsten bewertet, aber auch der "Freispruch für die Nicht-Tötung" erhält einen relativ hohen Nützlichkeitswert. Die Bestrafung für die

1) Es sei dem Leser überlassen, diese Werte in einer konkreten Situation zu begründen.

2) Die letzte Gleichung gilt, weil E_2 und E_3 voneinander unabhängig sind. Wäre das nicht der Fall, so hätte man die vorletzte Gleichung in der Zeile benutzen müssen. Das gleiche gilt für die nachfolgenden Formeln.

beiden anderen Tatbestände wird weniger hoch bewertet. Alle nicht tatschuldangemessenen, d.h. zu harten oder zu milden Strafen für die entsprechenden Tatbestände erhalten dagegen den Wert 0.

Aus dieser Nützlichkeitsmatrix gewinnen wir zusammen mit dem bereits abgeleiteten Wahrscheinlichkeitsvektor die folgenden Erwartungswerte der Maßnahmen:

$$E_u(M_1) = 0.2; E_u(M_2) = 0.1875; E_u(M_3) = 0.1; E_u(M_4) = 0.18.$$

Somit erhält "Bestrafung nach §211 StGB (wegen Mordes)" den höchsten Erwartungswert des Nutzens.

Vergleicht man die Ergebnisse der drei Beispiele, so kann man folgendes sehen: Während die Ableitung der Strafe nach §244 (wegen Diebstahls mit Waffen) als beste Maßnahme im ersten Beispiel akzeptabel erscheinen mag (denn sowohl die Wahrscheinlichkeit für "Diebstahl" wie auch für "Waffen" war sehr hoch (0.9)), so ist das Ergebnis in den beiden letzten Beispielen nicht befriedigend. Im zweiten Beispiel erhält sowohl "Bestrafung nach §250 StGB (wegen schweren Raubes)" wie auch "Bestrafung nach §244 StGB (wegen Diebstahls mit Waffen)" den gleichen höchsten Erwartungswert des Nutzens. Aus entscheidungstheoretischer Sicht kann daher (bei der endgültigen Wahl) jede dieser beiden Maßnahmen gewählt werden. Juristisch gesehen ist es aber ein Unterschied, ob man jemanden wegen "schweren Raubes" oder wegen "Diebstahls mit Waffen" verurteilt, denn im ersten Fall liegt die Freiheitsstrafe nicht unter fünf Jahren, im letzten Fall kann der Angeklagte sogar nur 6 Monate erhalten.

Im dritten Beispiel haben wir "Bestrafung nach §211 (wegen Mordes)" als die beste Entscheidung abgeleitet, obwohl der entsprechende Tatbestand "Mord" (zusammen mit "minder schwerer Fall von Totschlag") den geringsten Wahrscheinlichkeitswert hatte; außerdem war die Wahrscheinlichkeit für "Habgier" (die den Tatbestand "Mord" mitkonstituierte) 'nur' mit der Wahrscheinlichkeit von 0.5 festgestellt, d.h. das "Mord" qualifizierende Merkmal wurde nicht einmal mit der überwiegenden Wahrscheinlichkeit festgestellt. Wir müssen somit einen Weg finden, derartige Ergebnisse zu verhindern. Im übernächsten Paragraphen werden wir aus diesem Grund gewisse Einschränkungen formulieren. Vorher wollen wir uns aber mit dem Begriff des juristischen Nutzens auseinandersetzen.

§2. Der juristische Nutzen

In diesem Paragraphen werden wir uns ausführlich mit dem Begriff des juristischen Nutzens befassen. Der Nutzenbegriff ist neben dem der Wahrscheinlichkeit der zweite wichtigste Grundbegriff der Entscheidungstheorie. Somit ist seine Klärung im Rahmen einer juristischen Theorie wichtig. Bei dem juristischen Nutzen darf es sich offensichtlich nicht um einen persönlichen Nutzen handeln: denn erstens wären dann die juristischen Forderungen nach Gleichheit, Unparteilichkeit und Rechtssicherheit verletzt und zweitens wäre eine solche Beurteilung, die auf den persönlichen Nutzen des Richters abzielt, in ihrem Wesen ungerecht und anmaßend, denn die Interessen der eigentlich Betroffenen würden dabei zu wenig oder überhaupt nicht zur Geltung kommen.

Der juristische Nutzen muß somit unter dem sozialen Aspekt gesehen werden, der sowohl den Belangen der durch die Straftat Betroffenen wie auch denen der Gesellschaft und schließlich auch denen des Angeklagten gerecht wird. Der Richter muß sich also bei seiner subjektiven Nützlichkeitsbewertung der Konsequenzen fragen, wie nützlich bzw. wie wünschenswert die Folgen seiner Entscheidung für die Betroffenen (im weitesten Sinn) sein werden und nur dies bei seiner Nutzenzuordnung berücksichtigen. Es ist klar, daß auch dann noch genug Freiraum für eine unterschiedliche Bewertung der Konsequenzen bleibt, je nachdem, unter welchem Gesichtspunkt (genauer: unter welcher Zielvorstellung) die Entscheidung getroffen wird, z.B. demjenigen der Resozialisierung oder demjenigen der Wiedergutmachung des Übels usw., und je nachdem, wie diese Zielvorstellungen im konkreten Fall vom konkreten Individuum interpretiert werden.

Wie ich bereits anklingen ließ, erleichtert die Rechtstheorie unsere Aufgabe, indem sie ihrerseits bestimmte Auffassungen (strafrechtliche Theorien) von Sinn und Zweck der Strafe formuliert, in denen verschiedene juristische Zielsetzungen und -vorstellungen inhaltlich ausgelegt werden. Einige dieser Ziele seien hier erwähnt: Verbrechensbekämpfung und -verhütung, Resozialisierung (Wiedereingliederung in die Gesellschaft), Wiedergutmachung, Vergeltung, Genugtuung des Geschädigten, die Befriedigung des Sicherheitsbedürfnisses der Gesellschaft (etwa durch die Abschreckung potentieller Täter), angemessene Leidensverringerung des Bestraften und dergl.. Im folgenden werden wir uns jedoch nicht mit der juristischen (inhaltlichen) Interpretation all dieser Ziele befassen, sondern rein formal verschiedene Möglichkeiten untersuchen, diese Zielvorstellungen durch entsprechende Nützlichkeitsmatrizen zu repräsentieren. Auf diese Weise können wir auch besser bestimmen, in welcher Form und inwiefern derartige Vorstellungen einen positiven bzw. negativen Einfluß auf die Entscheidungsbildung ausüben.

Allerdings wird der Grundgedanke unserer Theorie des juristischen Nutzens eine (in der Rechtspraxis in dieser Form noch nicht formulierte) Gerechtigkeitsauffassung sein, die davon ausgeht, daß eine Straftat eine negative Veränderung des sozialen Gleichgewichts bzw. des juristischen Status Quo ist. Die Gerechtigkeit wird dabei rein formal im Sinne der gesetzlich vorgeschriebenen, also tatschuldan-

gemessenen Behandlung verstanden und als "formale Wiederherstellung des juristischen Status Quo" definiert. Alle übrigen strafrechtlichen Ziele werden auf dieses (fundamentale) Ziel "Gerechtigkeit" relativiert. Es wird außerdem gezeigt, daß nur die Gerechtigkeit als Ziel die Eindeutigkeit der Nutzenzuordnung garantiert (Wir werden eine sehr einfache, aber eindeutige Nützlichkeitsmatrix konstruieren, die wir sinngemäß als "Gerechtigkeitsmatrix" bezeichnen). Wegen der Eindeutigkeit der Nutzenzuordnung erfüllt Gerechtigkeit (als das einzige Ziel) die juristischen Forderungen nach der Gleichberhandlung vor dem Gesetz und (im weitesten Sinne) nach der Rechtssicherheit (als Berechenbarkeit und Kontrollierbarkeit des Rechts).

2.1. Eine informelle Betrachtung von Zielen

Wie bereits erwähnt, sind uns von der juristischen Seite bestimmte Auffassungen (strafrechtliche Theorien) über Sinn und Zweck der Strafe bekannt, in denen verschiedene juristische Zielsetzungen und -vorstellungen inhaltlich interpretiert werden. Juristisch relevante Ziele sind Verbrechensbekämpfung und -verhütung, Wiedergutmachung des Übels, Vergeltung (im Rahmen des gesetzlich Erlaubten), Genugtuung der Geschädigten, Resozialisierung (Wiedereingliederung und Besserung des Täters), die Befriedigung des Sicherheitsbedürfnisses der Gesellschaft (etwa durch Abschreckung), angemessene Leidensverringerung des Bestraften (wobei es sich bei den letzten drei um keine primären Ziele handeln dürfte¹⁾).

Verbrechensbekämpfung und -verhütung können ihrerseits offensichtlich auf verschiedene Weise erreicht werden: etwa durch Resozialisierung, Vergeltung oder aber auch durch Prävention (Allgemein- bzw. Einzelabschreckung).

Wiedergutmachung des Übels, Genugtuung der Geschädigten und die Befriedigung des Sicherheitsbedürfnisses der Gesellschaft können ihrerseits z.B. durch Vergeltung erreicht werden, die dritte aber

1) Möglicherweise wird die Bezeichnung der Resozialisierung und der Befriedigung des Sicherheitsbedürfnisses der Gesellschaft als sekundäre Ziele von manchem nicht akzeptiert. Darin primäre Ziele zu sehen, heißt aber zu vergessen, daß durch den Täter ein gewisses Leid bzw. Unrecht dem Opfer zugefügt wurde, welches dem Opfer gegenüber (ideell oder materiell) ausgeglichen werden muß, etwa, indem man dem Täter seinerseits ein (ideell oder materiell 'äquivalentes' Leid zufügt, bzw. indem der Täter auf sich ein 'äquivalentes' Leid nimmt).

auch durch die Abschreckung potentieller Täter (Generalprävention)¹⁾

Die gerechte Leidensverringerung des Bestraften kann etwa durch die Resozialisierung bewirkt werden.

Die strafrechtlichen Auffassungen bzw. Zielvorstellungen, die auf die Gestalt der Nützlichkeitsmatrix Einfluß ausüben, können wir grob gesagt in 4 Kategorien unterteilen, in: (1) resozialisierende, (2) vergeltende, (3) (general- bzw. spezial-) präventive und (4) wiederherstellende bzw. wiedergutmachende Vorstellungen.

Den obigen Zielen müssen ihrerseits offenbar demokratische Ziele wie Gerechtigkeit und Gleichbehandlung (= Gleichheit vor dem Gesetz) zugrunde liegen. Was unter Gleichbehandlung zu verstehen ist, dürfte, intuitiv gesehen, klar sein: Sie bedeutet, daß es bei der Bestrafung nicht auf individuelle Unterschiede wie Person, Geschlecht, Rasse, Angehörigkeit zu einer bestimmten Bevölkerungs- oder Berufsgruppe usw., ankommt, sondern einzig und allein auf die Verbrechenart. Später werden wir sehen, daß der Gleichbehandlungsgedanke im Widerspruch zu der Resozialisierungsauffassung stehen kann, in welcher individuelle Unterschiede den Urteilsspruch entscheidend prägen können.

Schwieriger ist die Interpretation des Gerechtigkeitsbegriffs. Wir werden darunter ein juristisch angemessenes Verhalten im Sinn der gerechten (d.h. gesetzlich gebotenen bzw. tatschuldangemessenen) Behandlung oder allgemeiner: die formale Wiederherstellung des juristischen Status Quo verstehen.²⁾ (Dazu mehr auf S.330ff dieser Arbeit).

1) Ob das auch in der Praxis empirisch belegt ist, soll hier dahin gestellt bleiben. Hier geht es zunächst um das theoretische Verständnis (intuitiver) juristischer Zielsetzungen.

2) Dieser Gedanke geht übrigens auf Aristoteles (Nik.Ethik,V) zurück. Die Gerechtigkeit (justitia commutativa) fordert nach Aristoteles u.a. z.B. die volle Wiedergutmachung bzw. Wiederherstellung des rechtswidrig zugefügten Schadens. Da die "volle Wiedergutmachung" bzw. die vollständige Wiederherstellung des Zustandes vor dem Begehen des Unrechts unmöglich ist, haben wir die Gerechtigkeit oben lediglich als formale Wiederherstellung des 'vorurrechtlichen' Zustandes (durch die dem Unrecht entsprechende Behandlung des Schuldigen) definiert.

2.2. Vergleich zwischen verschiedenen Nützlichkeitsmatrizen

Die beiden folgenden Nützlichkeitsmatrizen seien gegeben:

	Diebstahl (T)	kein Diebstahl (-T)
(1)		
M_1 : Strafe	1	0.3
M_2 : keine Strafe	0	0.6
(2)		
M'_1 : Strafe	0.5	0
M'_2 : keine Strafe	0.4	1

Interpretiert man diese Matrizen unter dem Gesichtspunkt der oben angesprochenen strafrechtlichen Zielsetzungen, so drückt die erste Matrix die Auffassung aus, daß es primär auf die tatschuldangemessene Bestrafung des Täters, d.h. auf Vergeltung für die begangene Straftat (ausgedrückt durch den Wert 1) ankommt und erst sekundär auf die gerechte Behandlung im Fall des "Nicht-Diebstahls"; aber auch der Gedanke der Abschreckung spielt offensichtlich eine Rolle, denn der "Strafe für keinen Diebstahl" ist auch ein positiver Nützlichkeitswert zugeordnet¹⁾. Es ist allerdings nicht klar, ob es dabei um die Allgemeinabschreckung potentieller Täter bzw. der Öffentlichkeit geht oder um die spezielle Abschreckung des Angeklagten (etwa weil er bereits in der Vergangenheit zahlreiche Straftaten begangen hat oder weil zu erwarten ist, daß er - auch wenn er die gegenwärtige Straftat nicht begangen hat - in der Zukunft irgendwelche Straftaten begehen wird).

Der zweiten Matrix liegt anscheinend ein 'milder' altruistischer Ansatz zugrunde, der nicht unbedingt auf Rache oder Vergeltung abzielt. Nach diesem Konzept entsteht der geringste Nutzen (= 0), wenn ein Nicht-Täter verurteilt wird, der größte (=1), wenn ein Nicht-Täter nicht bestraft wird. "Strafe für den Diebstahl" erfährt zwar eine (geringfügig) höhere Bewertung als "keine Strafe für den Diebstahl", ist jedoch bei weitem nicht so erstrebenswert wie die Verhütung dessen, daß ein Nicht-Täter zu Unrecht bestraft wird²⁾. Die zweite Matrix läßt sich nicht nur altruistisch, sondern

1) Vgl. S. 291, Fn.1 dieser Arbeit.

2) Um einem Mißverständnis vorzubeugen, sei hier klargestellt, daß es sich bei der obigen altruistischen Auffassung um keine in der Rechtspraxis etablierte Theorie über den Zweck und Sinn der Strafe handelt.

auch unter dem Gesichtspunkt der Resozialisierung interpretieren: So führt in diesem konkreten Fall offenbar die Bestrafung des Täters bzw. seine Nicht-Bestrafung zum fast gleichen (durchschnittlich hohen) resozialisierenden Erfolg, d.h. Bestrafung und Nichtbestrafung des Täters sind fast gleichermaßen gut für seine Resozialisierung¹⁾.

Der Wahrscheinlichkeitsvektor möge in beiden Fällen so aussehen:

$$[0.9, 0.1]$$

Der erwartete Nutzen der ersten Maßnahme im ersten Fall beträgt somit:

$$E_u(M_1) = 1 \cdot 0.9 + 0.3 \cdot 0.1 = 0.93;$$

der der zweiten:

$$E_u(M_2) = 0 \cdot 0.9 + 0.6 \cdot 0.1 = 0.06.$$

Also muß nach der Regel von Bayes die erste Maßnahme M_1 ("Strafe") vorgezogen werden.

Der erwartete Nutzen der ersten Maßnahme im zweiten Fall ist:

$$E_u(M'_1) = 0.5 \cdot 0.9 + 0 \cdot 0.1 = 0.45;$$

und der zweiten:

$$E_u(M'_2) = 0.4 \cdot 0.9 + 1 \cdot 0.1 = 0.46.$$

Also muß hier die zweite Maßnahme M'_2 ("keine Strafe") gewählt werden.

Mit der Wahrscheinlichkeit 0.91 für den Diebstahl ändert sich jedoch im zweiten Fall die Situation völlig. Der Erwartungswert von "Strafe" und "keine Strafe" wird dann gleich 0.455. Für $p > 0.91$ muß somit die erste Maßnahme M'_1 vorgezogen werden.

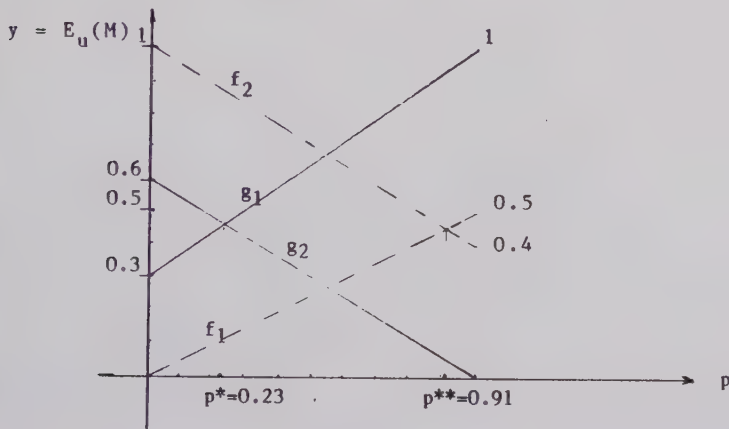
Es ist interessant festzustellen, daß im Fall der obigen Vergeltungsauffassung mit präventiver Wirkung die zweite Maßnahme erst dann vorgezogen wird, wenn die Wahrscheinlichkeit für den "Diebstahl" kleiner als 0.23 ist (im Unterschied zum zweiten Fall, wo sogar $P(T) = 0.9$ nicht genügt, um die Strafe zu wählen). Vgl. die nachfolgende Abbildung²⁾. Dieses Ergebnis entspricht auch der intuitiven Vorstellung, wonach im Mittelpunkt einer Vergeltungskonzeption mit präventiver Wirkung eben Vergeltung und Abschreckung stehen und daher auch eine geringere Wahrscheinlichkeit für das Be-

1) Im Unterschied zur (nichtjuristischen) 'altruistischen' Auffassung sind resozialisierende sozialpräventive Erwägungen heutzutage in der Rechtsprechung von großer Bedeutung.

2) Zur graphischen Lösung des Problems der Maximierung des Erwartungsnutzens vgl. S. 249f dieser Arbeit.

stehen des Sachverhalts in Kauf genommen wird; den Mittelpunkt einer altruistischen Konzeption bildet dagegen die Auffassung, daß kein Mensch zu Schaden kommen soll (daher die vorsichtigeren Wahl der härteren Maßnahme erst bei sehr hohen Wahrscheinlichkeiten für den Sachverhalt); im Mittelpunkt der Resozialisations-theorie steht die Wiedereingliederung des Täters in die Gesellschaft; diese wird u.U. erschwert oder verhindert durch die Verurteilung oder aber auch begünstigt durch die Nicht-Verurteilung des Täters.

Graphische Darstellung



$$g_1: y = 0.7p + 0.3; \quad g_2: y = -0.6p + 0.6; \quad p^* = 0.2\bar{3} \text{ (vgl. S. 250).}$$

$$f_1: y = 0.5p; \quad f_2: y = -0.6p + 1; \quad 0.5p^{**} = -0.6p^{**} + 1; \quad p^{**} = 0.9\bar{1}.$$

2.3. Formale Darstellung von Zielen

Bis jetzt sind wir davon ausgegangen, daß das Entscheidungssubjekt lediglich ein einziges Ziel oder höchstens zwei Ziele¹⁾ für seine Entscheidung in Betracht zieht. Es kann jedoch der Fall eintreten, wo mehr als eine bzw. mehr als zwei Zielvorstellungen den Ausgangspunkt einer Entscheidungssituation bilden. Dies wollen wir nun eingehend untersuchen.

Verschiedene Ziele können als miteinander konsistent oder aber als einander widersprechend empfunden werden. Das soll jedoch nicht so verstanden werden, daß der Entscheidende ein für alle Male für sich festgelegt hat, was für ihn verträgliche bzw. unverträgliche Ziele sind. Vielmehr kommt es hier allein auf die konkrete Entscheidungssituation an: Das Entscheidungssubjekt kann beispielsweise in einer Entscheidungssituation Vergeltung und Gerechtigkeit gleichzeitig anstreben, in einer anderen dagegen vor der ausschließlichen Wahl stehen: nur Vergeltung oder nur Gerechtigkeit. Wir wollen beide Möglichkeiten, d.h. einerseits verträgliche, andererseits unverträgliche Ziele, uns näher anschauen. Wir beginnen wiederum mit einem Beispiel. Die Entscheidungsprämissen mögen sein: "Diebstahl"/"kein Diebstahl" (Tatbestände) und "Strafe"/"keine Strafe" (Maßnahmen).

Weiter nehmen wir an, daß X zwei Ziele (die ihrerseits einem weiteren Ziel - etwa Verbrechenbekämpfung und -verhütung - untergeordnet sein können) in Betracht zieht: Vergeltung und/oder Resozialisation.

Die jeweiligen Nützlichkeitsmatrizen von X sehen z.B. wie folgt aus:

$$\text{Vergeltung: } \begin{bmatrix} 1 & 0 \\ 0 & 0.6 \end{bmatrix}$$

$$\text{Resozialisation: } \begin{bmatrix} 0.5 & 0 \\ 0.4 & 1 \end{bmatrix}$$

Die beiden Matrizen repräsentieren also eine subjektive Auffassung von X von Vergeltung und Resozialisation in diesem konkreten Fall.

Der Wahrscheinlichkeitsvektor habe in beiden Fällen die Gestalt:

$$[p_1, p_2]$$

p_1, p_2 sind beliebige Wahrscheinlichkeitswerte zwischen 0 und 1 mit $p_2 = 1 - p_1$.

1) Vgl. das Beispiel auf S.309, wo in einem Fall ein kombiniertes Ziel: Vergeltung mit präventiver Wirkung, d.h. also Vergeltung+Abschreckung, verfolgt wird.

Verträgliche Ziele

Angenommen, X hält Vergeltung und Resozialisierung für miteinander verträglich. Ferner will er beide Auffassungen bei seiner Entscheidungsbildung berücksichtigen, wobei er auf beide das gleiche Gewicht legt. Seinem Wunsch könnte offensichtlich entsprochen werden, indem man X vorschlägt, den subjektiven Wert jeder der beiden Maßnahmen aufgrund der Addition der beiden Nützlichkeitsmatrizen und anschließend der Multiplikation mit dem Wahrscheinlichkeitsvektor festzustellen und dann nach der Regel von Bayes zu wählen, d.h. es soll gelten:

$$M_i \succcurlyeq M_j \leftrightarrow \sum_{k=1}^n u_{ik} p_k + \sum_{k=1}^n u'_{ik} p_k \geq \sum_{k=1}^n u_{jk} p_k + \sum_{k=1}^n u'_{jk} p_k$$

$$\leftrightarrow E_u(M_i) + E'_u(M_i) \geq E_u(M_j) + E'_u(M_j)$$

$$\text{mit } E_u(M_i) = \sum_{k=1}^n u_{ik} p_k \quad \text{und} \quad E'_u(M_i) = \sum_{k=1}^n u'_{ik} p_k$$

bzw. in allgemeiner Form

(l sei die Anzahl der Nützlichkeitsmatrizen):

$$M_i \succcurlyeq M_j \leftrightarrow \sum_{k=1}^n (u_{ik}^1 + u_{ik}^2 + \dots + u_{ik}^l) \cdot p_k \geq$$

$$\sum_{k=1}^n (u_{jk}^1 + u_{jk}^2 + \dots + u_{jk}^l) \cdot p_k$$

$$\leftrightarrow \sum_{m=1}^l E_u^m(M_i) \geq \sum_{m=1}^l E_u^m(M_j)$$

$$\text{mit } E_u^m(M_i) = \sum_{k=1}^n u_{ik}^m \cdot p_k.$$

In unserem Fall wird sich X für "Strafe" bzw. "keine Strafe" entscheiden, je nachdem, ob

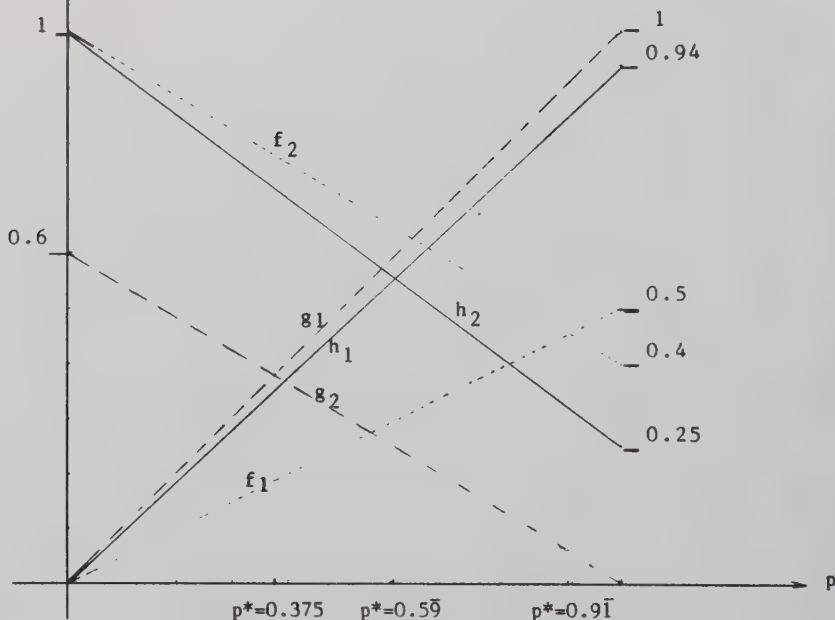
$$1 \cdot p_1 + 0 \cdot p_2 + 0.5 \cdot p_1 + 0 \cdot p_2 \text{ größer als}$$

$$0 \cdot p_1 + 0.6 \cdot p_2 + 0.4 \cdot p_1 + 1 \cdot p_2 \text{ ist oder umgekehrt.}$$

Zur Veranschaulichung betrachte man die folgende Abbildung und die Tabelle.

(Wir kürzen im folgenden die Vergeltungskonzeption mit VT, die Resozialisationskonzeption mit RT ab).

$$y = E_u(M)$$



$$g_1: y = 1 \cdot p + 0 \cdot (1-p) = p$$

$$g_2: y = 0 \cdot p + 0.6 \cdot (1-p) = -0.6 \cdot p + 0.6$$

$$p^* = -0.6 \cdot p^* + 0.6$$

$$p^* = 0.375$$

$$f_1: y = 0.5 \cdot p + 0 \cdot (1-p) = 0.5 \cdot p$$

$$f_2: y = 0.4 \cdot p + 1 \cdot (1-p) = -0.6 \cdot p + 1$$

$$0.5 \cdot p^* = -0.6 \cdot p^* + 1$$

$$p^* = 0.91$$

$$h_1: y = 0.94 \cdot p + 0 \cdot (1-p) = 0.94 \cdot p$$

$$h_2: y = 0.25 \cdot p + 1 \cdot (1-p) = -0.75 \cdot p + 1$$

$$0.94 \cdot p^* = -0.75 \cdot p^* + 1$$

$$p^* = 0.59$$

Tabelle

p		$E_u(M)$	$E'_u(M)$	$E_u(M) + E'_u(M)$
1	$E_u(M_1)$	1	0.5	1.5
	$E_u(M_2)$	0	0.4	0.4
0.99	$E_u(M_1)$	0.99	0.495	1.494
	$E_u(M_2)$	0.006	0.406	0.412
0.91	$E_u(M_1)$	0.91	0.455	1.365
	$E_u(M_2)$	0.054	0.455	0.509
0.9	$E_u(M_1)$	0.9	0.45	1.35
	$E_u(M_2)$	0.06	0.46	0.52
0.8	$E_u(M_1)$	0.8	0.4	1.2
	$E_u(M_2)$	0.12	0.52	0.64
0.7	$E_u(M_1)$	0.7	0.35	1.05
	$E_u(M_2)$	0.18	0.58	0.76
0.6	$E_u(M_1)$	0.6	0.3	0.9
	$E_u(M_2)$	0.24	0.64	0.88
0.59	$E_u(M_1)$	0.59	0.295	0.89
	$E_u(M_2)$	0.246	0.646	0.89
0.51	$E_u(M_1)$	0.51	0.255	0.765
	$E_u(M_2)$	0.294	0.694	0.988
0.5	$E_u(M_1)$	0.5	0.25	0.75
	$E_u(M_2)$	0.3	0.7	1
0.4	$E_u(M_1)$	0.4	0.2	0.6
	$E_u(M_2)$	0.36	0.76	1.12
0.3	$E_u(M_1)$	0.3	0.15	0.45
	$E_u(M_2)$	0.42	0.82	1.24
0.2	$E_u(M_1)$	0.2	0.1	0.3
	$E_u(M_2)$	0.48	0.88	1.36
0.1	$E_u(M_1)$	0.1	0.05	0.15
	$E_u(M_2)$	0.54	0.94	1.48
0	$E_u(M_1)$	0	0	0
	$E_u(M_2)$	0.6	1	1.6

" $E_u(M_i)$ " steht für den subjektiven Erwartungswert des Nutzens der i-ten Maßnahme, $i = 1, 2$. " $E_u(M)$ " ist der Erwartungswert nach der VT; " $E'_u(M)$ " ist ein solcher nach der RT. "p" ist die Wahrscheinlichkeit des Tatbestandes "Diebstahl".

Aus der Tabelle bzw. der Abbildung ersieht man, daß im Fall der VT die erste Maßnahme ("Strafe") bereits für $p > 0.375$ vorgezogen wird, im Fall der RT eine solche aber erst für $p > 0.91$. Falls hingegen das Entscheidungssubjekt die beiden Ziele gleichzeitig und im gleichen Maße verfolgt, wird es die erste Maßnahme für $p \in (0.59, 1]$ vorziehen. Dies stellt bzgl. des ersten Ziels eine Einschränkung und bzgl. des zweiten eine Erweiterung dar. Intuitiv besagt das: Wenn man Vergeltung und Resozialisation als gleichwertige Ziele ansieht und seine Nützlichkeitsmatrix dementsprechend konstruiert, so wird man aufgrund der zusammengesetzten Nützlichkeitsmatrix zu einer Auffassung gelangen, die 'milder' ist bzgl. der Vergeltung und 'strenger' bzgl. der Resozialisation. Wahrscheinlichkeitstheoretisch interpretiert bedeutet eine solche Auffassung, daß im Fall des ersten Ziels strengere Anforderungen an den Wahrscheinlichkeitswert für den Tatbestand gelten und mildere im Falle des zweiten.

Für $p > 0.91$ wird sowohl nach der VT als auch nach der RT und auch nach beiden zusammen die erste Maßnahme vorgezogen. In diesem Bereich stehen die beiden Ziele also nicht im Konflikt zueinander. Für $p \in (0.51, 0.91)$ wird nach der VT die erste Maßnahme vorgezogen, nach der RT die zweite, nach beiden zusammen jedoch die erste. Anschaulich gesprochen bedeutet das für die RT, daß sie in diesem Bereich zu streng bzgl. der Tatbestandsfeststellung ist. Für $p \in (0.375, 0.51)$ wird nach der VT die erste Maßnahme vorgezogen, nach der RT die zweite, nach beiden zusammen die zweite; d.h. die VT ist in diesem Bereich zu nachsichtig bzgl. der Tatbestandsfeststellung. Für $p \in [0, 0.375)$ sind die beiden Ziele wiederum miteinander verträglich: In beiden Fällen wird die zweite Maßnahme vorgezogen.

Im übrigen kann man die beiden Nützlichkeitsmatrizen auf S.312 durch eine äquivalente ersetzen, die transformiert so aussieht:

$$\begin{bmatrix} 0.94 & 0 \\ 0.25 & 1 \end{bmatrix}$$

Auch für sie gelten alle oben angestellten wahrscheinlichkeitstheoretischen Betrachtungen (vgl. die Abbildung).

Wir wollen die Entscheidungssituation, in der die Ziele miteinander konsistent sind, noch an einem weiteren Beispiel veranschaulichen. Es seien drei Tatbestände: "Mord", "Totschlag", "Notwehr" gegeben,

die X als Kandidaten für seine Entscheidung in Betracht zieht. Die entsprechenden Maßnahmen sind: "lebenslänglich", "nicht unter 5 Jahren" und "keine Strafe". Die Nützlichkeitsmatrizen mögen so aussehen:

	T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3
M_1 : lebenslänglich	$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 0.5 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 0.5 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 1 \\ 0.5 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$
M_2 : nicht unter 5 J.									
M_3 : keine Strafe									

Die obige Entscheidungssituation ist nicht ganz korrekt. Da es 3-teilweise qualifizierte-Tatbestände sind, gibt es insgesamt $2^3 = 8$ Ereignisse (vgl. S.297ff). Sei "T" die Abkürzung für "Tötung", "E" diejenige für Merkmale, die eine Tötung als Mord qualifizieren, "E*" diejenige für Merkmale, die eine Tötung als Notwehr qualifizieren. Die 8 Ereignisse sind: $T \wedge E \wedge E^*$, $T \wedge E \wedge \neg E^*$, $T \wedge \neg E \wedge E^*$, $T \wedge \neg E \wedge \neg E^*$, $\neg T \wedge E \wedge E^*$, $\neg T \wedge E \wedge \neg E^*$, $\neg T \wedge \neg E \wedge E^*$, $\neg T \wedge \neg E \wedge \neg E^*$. Wir können juristische Tatbestände wiederum (wie auf S.297ff) mit folgenden Ereignissen identifizieren: "Mord" $\hat{=}$ " $T \wedge E \wedge E^*$ ", "Totschlag" $\hat{=}$ " $T \wedge \neg E \wedge E^*$ ", "Notwehr" $\hat{=}$ " $((T \wedge \neg E \wedge E^*) \vee (T \wedge E \wedge \neg E^*))$ " = " $T \wedge E^*$ ". Falls $P(T) = 1$, reduziert sich die Situation auf drei Ereignisse $T \wedge E \wedge \neg E^* = E \wedge \neg E^*$, $T \wedge \neg E \wedge \neg E^* = \neg E \wedge \neg E^*$ und $T \wedge E^* = E^*$. "Mord" entspricht dann " $E \wedge \neg E^*$ ", "Totschlag" $\hat{=}$ " $\neg E \wedge \neg E^*$ " und "Notwehr" $\hat{=}$ " E^* ". So gesehen sind die obigen Tatbestände T_1 ("Mord"), T_2 ("Totschlag") und T_3 ("Notwehr") disjunkt und erschöpfend (wobei gilt: $P(\text{Tötung}) = 1$; würde das nicht gelten, so wären diese drei Tatbestände nicht erschöpfend).

Die erste der obigen Nützlichkeitsmatrizen repräsentiere die Auffassung von X von Gerechtigkeit (denn darin bekommen alle tatschuldangemessenen, also aus juristischer Sicht gerechten, Beurteilungen für entsprechende Tatbestände den höchsten Nützlichkeitswert ($=1$), alle nicht tatschuldangemessenen Beurteilungen den niedrigsten Wert ($=0$)). Die beiden anderen (die sich voneinander nur in der zweiten Spalte unterscheiden) mögen die Auffassung von X bzgl. der Resozialisierung repräsentieren: Nach der zweiten Matrix zieht er im Fall von "Mord" eine geringere Bestrafung dem Freispruch vor; nach der dritten zieht er außerdem im Fall von "Totschlag" den Freispruch der lebenslangen Strafe vor (X schwankt also in diesem konkreten Fall zwischen zwei divergierenden Interpretationen der Resozialisierung). Es sei angenommen, daß die drei Nützlichkeitsmatrizen für X gleichwertig und miteinander verträglich sind. Man kann sie somit in eine gemeinsame Nützlichkeitsmatrix überführen, die transformiert so aussieht:

$\begin{bmatrix} 1 & 0 & 0 \\ 1/3 & 1 & 0 \\ 0 & 1/6 & 1 \end{bmatrix}$

Gewichtung von Zielen

Falls X mehrere Ziele, jedoch nicht im gleichen Verhältnis, berücksichtigen möchte, so kann er durch die Gewichtung einzelner Ziele dieser Intention Ausdruck verleihen.

Ist im Beispiel auf S.312 das Verhältnis der Ziele: Resozialisierung (Z_1) und Vergeltung (Z_2) 3 zu 1, d.h. $Z_1 = 3Z_2$, so hat die gemeinsame Nützlichkeitsmatrix dementsprechend die Gestalt:

$$\begin{bmatrix} 0.69 & 0 \\ 0.33 & 1 \end{bmatrix}$$

Man sieht eine deutliche Verschiebung zugunsten der RT, denn für $p < 0.74$ wird die zweite und erst für $p > 0.74$ die erste Maßnahme vorgezogen.

Allgemein gilt für die Konstruktion einer mehreren Zielvorstellungen Rechnung tragenden Nützlichkeitsmatrix aus den Nützlichkeitsmatrizen, in denen jeweils ein Ziel zum Ausdruck gebracht wird: Man ordnet jedem Nützlichkeitswert in der jeweiligen Nützlichkeitsmatrix eine reelle Zahl (sein Gewicht¹⁾) zu, addiert anschließend die entsprechenden Nützlichkeiten in allen Matrizen zueinander und normiert schließlich die so erhaltene Nützlichkeitsmatrix.

Formal sieht diese Operation so aus:

Es seien k Nützlichkeitsmatrizen gegeben, in denen jeweils einzelne unterschiedliche Zielvorstellungen zum Ausdruck gebracht werden. Diese sind:

$$[u_{ij}^1], [u_{ij}^2], \dots, [u_{ij}^k], \quad i = 1, \dots, r, \quad j = 1, \dots, n.$$

Weiter sei angenommen, daß diesen Matrizen k (unterschiedliche oder gleiche) Gewichte, entsprechend der Bedeutung der jeweiligen Zielvorstellung, zukommen, welche durch reelle Zahlen $q_1, q_2, \dots, q_k$, repräsentiert werden. Es gilt also:

$$q_1[u_{ij}^1], q_2[u_{ij}^2], \dots, q_k[u_{ij}^k], \quad i = 1, \dots, r, \quad j = 1, \dots, n.$$

Aus diesen Nützlichkeitsmatrizen erhält man durch Addition die folgende Nützlichkeitsmatrix:

1) wobei natürlich jedem Nützlichkeitswert das gleiche Gewicht in der Nützlichkeitsmatrix zugeordnet werden muß.

$$\left[\sum_{m=1}^k q_m \cdot u_{ij}^m \right], \quad m = 1, \dots, k, \quad i = 1, \dots, r, \quad j = 1, \dots, n.$$

Sie enthält also alle Ziele bzw. Zielvorstellungen des Entscheidungssubjekts entsprechend ihrer Bedeutung (Gewichtung) und kann nun noch auf eine normierte Form gebracht werden.

Unverträgliche Ziele

Angenommen, X hält Vergeltung und Resozialisation für insofern miteinander unverträglich, als er seine Entscheidung unter klaren Zielvorstellungen fällen und deswegen keinen Kompromiß diesbezüglich dulden möchte. Da ihm aber die beiden Ziele gleichermaßen attraktiv erscheinen, gerät er nun in einen Gewissenskonflikt. Mag in der Praxis ein solcher Fall selten auftreten – ein Entscheidender wird in der Regel wohl wissen bzw. zu wissen glauben, welchen Zweck er mit der Wahl seiner Entscheidung verfolgt, und wenn es sich um mehrere Ziele handelt, sie auch allesamt (entsprechend ihrer Gewichtung) berücksichtigen wollen –, so möchten wir hier diese Möglichkeit nicht von vornherein ausschließen.

Eine patente Lösung haben wir für diesen Fall nicht; eine Entscheidungsregel könnte vielleicht aber trotzdem aufgestellt werden. Nach dieser Regel wird vom Entscheidenden erwartet, daß er den gesamten Erwartungsnutzen aller möglichen Maßnahmen relativ auf ein Ziel ermittelt und dann diesen Wert mit einem, der relativ auf ein anderes Ziel ermittelt wurde, vergleicht. Das Entscheidungssubjekt soll dann diejenige Zielvorstellung wählen, die ihm insgesamt den größten Erwartungswert liefert. Ein derartiger juristischer Opportunismus hätte sein Gegenstück etwa in der folgenden alltäglichen Situation: X will nach München und steht vor der Wahl, entweder mit dem Flugzeug zu fliegen (h_1) oder mit dem Zug zu fahren (h_2), wobei es ihm auf die Umstände "schönes Wetter" (U_1) und "schlechtes Wetter" (U_2) ankommt. Außerdem zieht er zwei in seinen Augen unverträgliche Ziele in Betracht: entweder die Reise zu genießen oder aber möglichst schnell nach München zu kommen, um dort ein auf ihn wartendes Geschäft zum Abschluß zu bringen. (Stillschweigend nehmen wir an, daß X nicht darauf angewiesen ist, das Geschäft möglichst schnell zu machen¹). Wir gehen also davon aus, daß X sich für das Ziel "Genießen

¹) Eine weitere stillschweigende Annahme ist, daß X nicht nur aus Geschäftsgründen nach München will.

der Reise" bzw. für das "schnelle Abwickeln eines Geschäfts" entscheiden will, je nachdem, wie die Entscheidungssituation für ihn insgesamt aussieht.

Seine Wahrscheinlichkeiten für "schönes Wetter" und "schlechtes Wetter" seien jeweils 0.9 und 0.1. Ferner stelle X die folgenden Nützlichkeitsmatrizen auf:

Genießen der Reise:	U_1	U_2	Geschäfts- abschluß	U_1	U_2
h_1	0.4	0.2	h_1	1	0
h_2	1	0	h_2	0.3	0.3

(Die erste Nützlichkeitsmatrix kann dahingehend interpretiert werden, daß X nicht nur schönes Wetter liebt und gerne die Landschaft bei schönem Wetter genießt, sondern auch ein wenig risikofreudig ist; dem Zugfahren bei schlechtem Wetter kann er hingegen nichts außer Langweile abgewinnen).

X wird also, anhand seiner subjektiven Bewertung der Situation, das Ziel "Genießen der Reise" vorziehen, weil diese Zielsetzung insgesamt für ihn mit einem größeren Nutzen verbunden ist (1.28 gegenüber 1.2).

Die obige Entscheidungssituation kann man übrigens auch im Grundmodell rekonstruieren, indem man "Genießen der Reise während man mit dem Flugzeug fliegt oder mit dem Zug fährt" (h'_1) und "Schnelles Abwickeln des Geschäfts, wobei man entweder mit dem Flugzeug fliegt oder mit dem Zug fährt" (h'_2) als die beiden möglichen Entscheidungsalternativen nimmt und die beiden alten Umstände beläßt; wir haben dann:

	U_1	U_2	
h'_1	$1+0.4$	$0+0.2$	$\begin{bmatrix} 1.4 & 0.2 \\ 1.3 & 0.3 \end{bmatrix}$
h'_2	$1+0.3$	$0+0.3$	

(Die obige Addition bei der Bewertung der Konsequenzen ergibt sich daraus, daß die Konsequenzen z.B. "Genießen der Reise während man mit dem Flugzeug fliegt oder mit dem Zug fährt bei schönem Wetter" nichts anderes ist als die zusammengesetzte Konsequenz aus "mit dem Flugzeug fliegen bei schönem Wetter" und "mit dem Zug fahren bei schönem Wetter" unter dem Aspekt "Genießen der Reise". Auch die anderen Konsequenzen sind als auf diese Weise gebildet zu denken).

Die Multiplikation von 1.4 und 0.2 (bzw. 1.3 und 0.2) mit den Wahrscheinlichkeiten 0.9 und 0.1 und die anschließende Summation ergibt für jede Entscheidung (Zielsetzung) nichts anderes als den früher errechneten Gesamterwartungswert des Nutzens einer Entscheidungssituation unter dem Gesichtspunkt der jeweiligen Zielsetzung (denn $E_U(h'_1) = 1.28$ und $E_U(h'_2) = 1.2$).

Angewandt auf eine juristische Entscheidungssituation kann man juristische Zielsetzungen auch als Entscheidungsalternativen in Betracht ziehen. Im Beispiel auf S.312 sind es dann "Vergeltung durch Verurteilung oder Freispruch" (M_1^1) und "Resozialisation durch Verurteilung oder Freispruch" (M_2^1). Mit den beiden Tatbeständen "Diebstahl" und "kein Diebstahl" erhalten wir vier Konsequenzen: "Vergeltung (bzw. Resozialisation) durch Verurteilung oder Freispruch für den Diebstahl", "Vergeltung (bzw. Resozialisation) durch Verurteilung oder Freispruch für keinen Diebstahl". Die Nützlichkeitsmatrix sieht dementsprechend so aus:

$$\begin{array}{cc} & T_1 & T_2 \\ \begin{array}{c} M_1^1 \\ M_2^1 \end{array} & \begin{bmatrix} 1 + 0 & 0 + 0.6 \\ 0.5 + 0.4 & 0 + 1 \end{bmatrix} & = \begin{bmatrix} 1 & 0.6 \\ 0.9 & 1 \end{bmatrix} \end{array}$$

Für $P(T_1) = 0.9$ erhalten wir die folgenden Erwartungswerte:

$E_u(M_1^1) = 0.96$ und $E_u(M_2^1) = 0.91$. Somit muß die erste Zielsetzung "Vergeltung" gewählt werden.

Wir können also die folgende Entscheidungsregel formulieren:

Bezeichnet Z_i das i -te Ziel, $E_u^i(M)$ den gesamten Erwartungswert des Nutzens relativ auf das i -te Ziel, $E_{u_i}(M_k)$ den Erwartungswert der k -ten Maßnahme relativ auf das i -te Ziel, so lautet die Regel:

$$Z_i \succ Z_j \leftrightarrow E_u^i(M) \geq E_u^j(M) \leftrightarrow \sum_{k=1}^r E_{u_i}(M_k) \geq \sum_{k=1}^r E_{u_j}(M_k).$$

$$\text{Da } E_u(M_k) = \sum_{l=1}^n u_{kl} p_l \quad \text{und} \quad \sum_{k=1}^r E_{u_i}(M_k) = \sum_{k=1}^r \sum_{l=1}^n u_{kl} p_l \quad \text{ist,}$$

können wir die Regel auch so schreiben (wobei die oberen Indizes angeben, relativ auf welches Ziel der gesamte Erwartungswert des Nutzens gilt):

$$Z_i \succ Z_j \leftrightarrow \sum_{k=1}^r \sum_{l=1}^n u_{kl}^i p_l \geq \sum_{k=1}^r \sum_{l=1}^n u_{kl}^j p_l.$$

Nach dieser Regel muß der Entscheidende diejenige Zielsetzung wählen, die ihm den größten gesamten Erwartungsnutzen garantiert.

Aus der Tabelle auf S.315 ersieht man, daß für $p \in (0.8, 1]$ der gesamte Erwartungswert des Nutzens relativ auf die VT höher ist als ein solcher relativ auf die RT. Also wird X, falls er nach dieser Regel handelt, die VT für $p \in (0.8, 1]$, die RT für $p \in [0, 0.8)$ wählen. Das ist folgendermaßen zu interpretieren: Da es dem Entscheidungssubjekt jetzt nicht auf einen Kompromiß zwischen verschiedenen

Zielsetzungen, sondern auf den höchsten gesamten Erwartungswert ankommt, muß es dementsprechend derjenigen Auffassung den Vorzug geben, die ihm einen solchen für einen bestimmten Wahrscheinlichkeitswert garantiert. Für $p = 0.8$ sind die beiden Auffassung für X gleichwertig. In diesem Fall muß X nach anderen (nicht entscheidungstheoretischen) Lösungen - etwa nach dem Prinzip der Humanität - Ausschau halten.

Eine andere nicht entscheidungstheoretische Lösung des obigen Zielkonflikts (die weniger opportun ist) bestünde z.B. im Wurf einer Münze, wenn es sich um zwei unverträgliche, jedoch gleichwertige Zielvorstellungen handelt, oder z.B. im Werfen einer "zwei" oder einer "drei" mit einem Würfel (wobei "zwei" und "drei" dem Ziel "Vergeltung" und die anderen Augenzahlen dem Ziel "Resozialisation" zugeordnet werden), falls beide Auffassungen ("Vergeltung" und "Resozialisation") nicht gleichwertig sind, sondern was ihre Gewichte angeht, im Verhältnis 1 zu 3 stehen.

2.4. Verallgemeinerung verschiedener Zielvorstellungen. Untersuchung an einem Beispiel

Betrachten wir die folgenden Gesetzesbestimmungen:

§211 StGB. Mord. (1) Der Mörder wird mit lebenslanger Freiheitsstrafe bestraft. (2) Mörder ist, wer aus Mordlust, zur Befriedigung des Geschlechtstriebes, ... einen Menschen tötet. §212 StGB. Totschlag.

(1) Wer einen Menschen tötet, ohne Mörder zu sein, wird als Totschläger mit Freiheitsstrafe nicht unter fünf Jahren bestraft. §32 StGB. Notwehr. Wer eine Tat begeht, die durch Notwehr geboten ist, handelt nicht rechtswidrig. (d.h. er wird freigesprochen).

Entscheidungstheoretisch rekonstruiert lautet somit die Konsequenzenmatrix¹⁾

	Mord	Totschlag	Notwehr
	T_1	T_2	T_3
M_1 : lebenslänglich	(M_1, T_1)	(M_1, T_2)	(M_1, T_3)
M_2 : nicht unter 5 Jahren	(M_2, T_1)	(M_2, T_2)	(M_2, T_3)
M_3 : keine Strafe	(M_3, T_1)	(M_3, T_2)	(M_3, T_3)

1) Dieser Fall ist ausführlicher auf S.316f dieser Arbeit beschrieben.

Im folgenden setzen wir voraus, daß die Nützlichkeitsmatrix durch die obigen Gesetzesbestimmungen teilweise determiniert ist, d.h. es wird hier der dem Tatbestand tatschuldangemessenen Maßnahme der Nutzenwert 1 zugeordnet bzw. die den Tatbeständen tatschuldangemessenen Maßnahmen durch den Nützlichkeitsvektor $(1,1,1)$ repräsentiert. (Eine solche "Determiniertheit" ist zwar juristisch verankert¹⁾, wird in der Praxis allerdings nicht immer gegeben sein: So kann ein Richter in einem konkreten Fall eine mildere tatschuldangemessene Strafe für einen leichteren Tatbestand für angebrachter halten als eine härtere tatschuldangemessene Strafe für einen schwereren Tatbestand oder umgekehrt und dementsprechend ihr einen höheren/geringeren Nutzenwert zuordnen. (Mit der obigen Determiniertheit soll dem Gedanken Ausdruck verliehen werden, daß, unabhängig von der jeweiligen Auffassung, die formale Gerechtigkeitsvorstellung - im Sinne der Gesetzesangemessenheit - in die Bestimmung der Gestalt einer Nützlichkeitsmatrix Eingang finden kann). Unter dieser Voraussetzung sieht die Nützlichkeitsmatrix so aus:

$$\begin{bmatrix} 1 & a_{12} & a_{13} \\ a_{21} & 1 & a_{23} \\ a_{31} & a_{32} & 1 \end{bmatrix} \quad \text{mit } a_{ij} \text{ als Variablen, } i \neq j$$

Wir können natürlich diese Matrix verallgemeinern. Seien $T_1, \dots, T_n$ Tatbestände und $M_1, \dots, M_n$ ²⁾ die diesen Tatbeständen entsprechenden Maßnahmen. Dann hat die obige Nützlichkeitsmatrix die folgende Gestalt:

$$\begin{bmatrix} 1 & a_{12} & \dots & a_{1n} \\ a_{21} & 1 & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & 1 \end{bmatrix}$$

1) Eben durch die Darstellung von Straftatenn mit den auf sie folgenden Strafen im Strafgesetzbuch. Dabei ist anzunehmen, daß aus gesetzgeberischer Sicht die durch Strafen für entsprechende Tatbestände angestrebten Zustände gleichwertig sind, d.h. es ist nicht der Fall, daß z.B. die lebenslange Freiheitsstrafe für den Mord als wünschenswerter oder verwerflicher angesehen wird als eine Gefängnisstrafe von "bis zu fünf Jahren" oder "Geldstrafe" für Diebstahl. Der gleiche Gedanke kommt auch im juristischen Syllogismus zum Ausdruck: Darin folgt auf einen Straftatbestand juristisch notwendig eine entsprechende (im StGB angegebene) Strafe.

2) Wir beschränken uns auf den Fall, wo jedem Tatbestand jeweils nur eine Maßnahme zukommt.

Wir kehren nun zu unserer 3x3-Matrix zurück. Mit dieser zum Teil determinierten Nützlichkeitsmatrix sind z.B. die folgenden konkreten Nützlichkeitsmatrizen vereinbar:

$$(a) \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Hier wird eine Strafe, die den Tatbestandsgehalt überschreitet, für gleich schlecht gehalten wie eine Strafe, die einen solchen unterschreitet. Wir können diese Vorstellung mit der auf S.308 erwähnten Gerechtigkeitsauffassung in Einklang bringen, indem wir die obige Nützlichkeitsmatrix wie folgt interpretieren: Man geht hier erstens von der Konzeption aus, wonach der gerechte Ausgleich des Übels im Mittelpunkt steht (ausgedrückt durch den Nützlichkeitswert 1 für die Konsequenzen, die die tatschuldangemessene (gesetzlich gebotene) Behandlung für die entsprechenden Tatbestände darstellen und 0 für die Konsequenzen "keine Strafe für Mord/Totschlag" und für eine mildere (also nicht tatschuldangemessene) Bestrafung des Mörders bzw. eine härtere Bestrafung (wegen Mordes), obwohl nur Totschlag vorliegt); zweitens setzt man voraus, daß nur der Schuldige für den Ausgleich des Übels in Frage kommt (das wird durch die Werte 0 für die Bestrafung des in Notwehr Handelnden und 1 für dessen Nicht-Bestrafung ausgedrückt); und drittens nimmt man an, daß die Bewertung der tatschuldangemessenen Konsequenzen weder von einem konkreten Tatbestand noch von der Art der Bestrafung abhängt, sondern einzig und allein von dem gesetzlich Gebotenen (ausgedrückt durch gleiche Nützlichkeitswerte): Demnach ist "lebenslange Strafe für Mord" ebenso wichtig wie "Freiheitsstrafe von 'nicht unter 5 Jahren' für Totschlag" und "Freispruch für Notwehr" (Es gibt also keinen Grund, eine Art von tatschuldangemessener Behandlung gegenüber den anderen vorzuziehen). Intuitiv gesehen, deckt sich eine solche Interpretation etwa damit, was man sich gewöhnlich unter der Gerechtigkeit vorstellt: der angemessene Ausgleich des Übels ist ebenso wichtig wie die Vermeidung dessen, daß ein Unschuldiger zu Schaden kommt¹⁾.

¹⁾Wir sind, wie bereits erwähnt, allerdings nicht an der inhaltlichen Interpretation des Gerechtigkeitsbegriffs als solchen interessiert, sondern an einer bestimmten Auffassung, die die Gerechtigkeit im Sinne der formalen Wiederherstellung des juristischen Status Quo definiert. Was darunter zu verstehen ist, wurde gerade am Beispiel der obigen Nützlichkeitsmatrix erläutert. Mehr zum Gerechtigkeitsbegriff siehe S.326f und S.330ff dieser Arbeit.

$$(b) \begin{bmatrix} 1 & 0 & 0 \\ x & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad 0 < x < 1$$

Hier wird eine geringere als die tatschuldangemessene Strafe für "Mord" für besser gehalten als gar keine, ansonsten ist alles wie bei der ersten Matrix. (Diese Matrix kann als zum Teil resozialisierend interpretiert werden).

$$(c) \begin{bmatrix} 1 & 0 & 0 \\ x & 1 & 0 \\ 0 & y & 1 \end{bmatrix} \quad 0 < x, y < 1$$

Hier wird "Freispruch für Totschlag" für besser gehalten als eine härtere Bestrafung für denselben Tatbestand, ansonsten ist alles wie bei Matrix (b). (Auch diese Auffassung kann als zum Teil resozialisierend interpretiert werden).

$$(d) \begin{bmatrix} 1 & y & 0 \\ x & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad 0 < x, y < 1$$

Hier wird eine härtere als die durch die Strafbestimmung vorgesehene Bestrafung für "Totschlag" für besser gehalten als gar keine, ansonsten ist alles wie bei Matrix (b). (Diese Auffassung kann als zum Teil präventiv interpretiert werden).

Seien im folgenden x und y in (b), (c) und (d) gleich 0.5. Wir wollen in einer Tabelle angeben, welche Maßnahme für welche Wahrscheinlichkeitswerte der Tatbestände jeweils aufgrund der Nützlichkeitsmatrizen (a), (b), (c) und (d) gewählt werden muß.

Matrix	vorgezogene Maßnahme	Wahrscheinlichkeitsverteilung für die Tatbestände ¹⁾
(a)	M_1	$P(T_1) > 0.5$, $P(T_2)$ und $P(T_3)$ beliebig;
	M_1	$0.34 \leq P(T_1) \leq 0.5$, $P(T_2) < P(T_1)$ und $P(T_3) < P(T_1)$;
	M_2	$P(T_2) > 0.5$, $P(T_1)$ und $P(T_3)$ beliebig;
	M_2	$0.34 \leq P(T_2) \leq 0.5$, $P(T_1) < P(T_2)$ und $P(T_3) < P(T_2)$;
	M_3	$P(T_3) > 0.5$, $P(T_1)$ und $P(T_2)$ beliebig;
	M_3	$0.34 \leq P(T_3) \leq 0.5$, $P(T_1) < P(T_3)$ und $P(T_2) < P(T_3)$.

1) beansprucht nicht, vollständig zu sein. Außerdem: für kein T gilt, daß $P(T) = 0$ ist. Wahrscheinlichkeitswerte werden im folgenden nur auf zwei Stellen nach dem Komma berücksichtigt.

- (b)
- | | |
|-------|--|
| M_1 | $P(T_1) \geq 0.7;$ |
| M_1 | $P(T_1) \geq 0.41, P(T_2) < 1/2P(T_1), P(T_3) < P(T_1);$ |
| M_2 | $0.5 \leq P(T_1) \leq 0.66, P(T_2) > 1/2P(T_1);$ |
| M_2 | $P(T_2) > 0.5, P(T_1) \text{ und } P(T_3) \text{ beliebig};$ |
| M_2 | $0.26 \leq P(T_2) \leq 0.5, P(T_3) < P(T_2) \text{ und } P(T_1) \leq 0.5;$ |
| M_3 | $P(T_3) > 0.5, P(T_1) \text{ und } P(T_2) \text{ beliebig.}$ |
- (c)
- | | |
|-------|---|
| M_1 | $P(T_1) \geq 0.7;$ |
| M_1 | $P(T_1) > 0.5, P(T_2) < 1/2P(T_1);$ |
| M_1 | $0.45 \leq P(T_1) \leq 0.5, P(T_2) < 1/2P(T_1) \text{ und } P(T_3) < P(T_1);$ |
| M_2 | $P(T_2) \geq 0.7;$ |
| M_2 | $P(T_2) > 0.5, P(T_3) < 1/2P(T_2);$ |
| M_2 | $0.35 \leq P(T_2) \leq 0.5, P(T_3) < 1/2P(T_2);$ |
| M_3 | $P(T_3) > 0.5;$ |
| M_3 | $0.34 \leq P(T_3) \leq 0.5, P(T_2) < P(T_3) \text{ und } P(T_1) < P(T_3).$ |
- (d)
- | | |
|-------|---|
| M_1 | $P(T_1) > 0.5;$ |
| M_1 | $0.29 < P(T_1) \leq 0.5, P(T_2) < P(T_1) \text{ und } P(T_3) < 0.43;$ |
| M_2 | $P(T_2) > 0.5;$ |
| M_2 | $0.29 < P(T_2) \leq 0.5, P(T_1) < P(T_2) \text{ und } P(T_3) < 0.43;$ |
| M_3 | $P(T_3) > 0.5.$ |

Aus der Tabelle ersieht man, daß ohne Einschränkung bzgl. der Wahrscheinlichkeitswerte für die Tatbestände alle Matrizen zu wenig befriedigenden Ergebnissen führen (die Wahl einer Maßnahme als der besten trotz relativ geringen (< 0.5) Wahrscheinlichkeiten für die entsprechenden Tatbestände); nur im Fall der Matrix (a) wird die Maßnahme gewählt, die dem überwiegenden (d.h. dem wahrscheinlicheren) Tatbestand entspricht, in (b), (c) und (d) dagegen in vielen Fällen eine Maßnahme, deren zugehöriger Tatbestand nicht mit überwiegender Wahrscheinlichkeit nachgewiesen ist.

Wir wollen nun die Nützlichkeitsmatrizen (a), (b), (c) und (d) in allgemeiner Form als $n \times n$ -Matrizen anschreiben.

Die Verallgemeinerung von (a) hat die Form:

$$GM \quad \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

In dieser Matrix werden die tatschuldangemessenen (gesetzlich gebotenen) Maßnahmen für die entsprechenden Tatbestände durch den Nutzenvektor $(1, \dots, 1)$ repräsentiert; dieser Vektor ist mit der Hauptdiagonale von links nach rechts identisch und hat n Komponenten. Den Maßnahmen, deren Schwere nicht den Tatbeständen entspricht,

sei es, daß sie sie übersteigen, sei es, daß sie sie nicht erreichen, werden Null-Nutzenwerte zugeordnet.

Wir wollen eine Konzeption, die von derartigen Vorstellungen ausgeht, allgemein als Gerechtigkeitskonzeption (kurz "GT") bezeichnen. Die Nützlichkeitsmatrix, die der Gerechtigkeitskonzeption zugrunde liegt, hat somit eine bestimmte Gestalt: Sie wird durch den Nützlichkeitsvektor $(1, \dots, 1)$ auf der Hauptdiagonale charakterisiert. Eine derartige Nützlichkeitsmatrix bezeichnen wir als Gerechtigkeitsmatrix (kurz "GM"). Auf S.330ff werden wir die Rechtfertigung für die hier vertretene Gerechtigkeitsauffassung und ihre formale Repräsentation durch die Matrix GM liefern.

Im Unterschied zu (a) kann man (b), (c) und (d) jeweils auf verschiedene Weise verallgemeinern. Wir wollen hier nur einige Beispiele angeben.

Z.B. kann eine Verallgemeinerung von (b) folgende Form haben:

$$(b_1) \quad \begin{bmatrix} 1 & 0 & \dots & 0 \\ a_{21} & 1 & \dots & 0 \\ a_{31} & a_{32} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{(n-1)1} & \dots & \dots & 0 \\ a_{n1} & a_{n2} & \dots & 1 \end{bmatrix} \quad \text{mit } a_{ij} \begin{cases} \neq 0, & \text{für } i \geq j \\ = 0, & \text{für } i < j \end{cases}$$

und $a_{ij} < a_{i-1j}, \quad i = 1, \dots, n, \quad j < i$

Aber auch so:

$$(b_2) \quad \begin{bmatrix} 1 & 0 & \dots & 0 \\ a_{21} & 1 & \dots & 0 \\ 0 & a_{32} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} \quad \text{mit } a_{ij} \begin{cases} \neq 0, & \text{für } i = j+1 \\ & i = j \\ = 0, & \text{für } i \neq j+1, \\ & i \neq j \end{cases}$$

Eine Verallgemeinerung von (c) kann so aussehen:

$$(c_1) \quad \begin{bmatrix} 1 & 0 & \dots & 0 \\ a_{21} & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a_{n2} & \dots & 1 \end{bmatrix} \quad \text{mit } a_{ij} \begin{cases} \neq 0, & \text{für } i \geq j, \\ & a_{ij} \neq a_{n1} \\ = 0, & \text{für } i < j \end{cases}$$

Aber auch wie (b_1) .

Als Verallgemeinerung von (d) kann z.B. die folgende Matrix angegeben werden:

$$(d_i) \quad \begin{bmatrix} 1 & a_{12} & \dots & 0 \\ a_{21} & 1 & \dots & 0 \\ 0 & a_{32} & \dots & 0 \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

$$\text{mit } a_{ij} \quad \begin{cases} \neq 0, \text{ für } i = j+1, \\ \quad \quad \quad j = i+1, i = j \\ \\ = 0, \text{ für } i \neq j+1, \\ \quad \quad \quad j \neq i+1, i \neq j \end{cases}$$

Eine solche Matrix besagt: Solange die Möglichkeiten bestehen, daß die Bestrafung milder ausfällt als es dem Tatbestand entspricht, oder härter, als es der Tatbestand vorsieht, müssen diese beiden Möglichkeiten einander die Waage halten.

Aus der obigen Schilderung ersieht man, daß nur die Matrix GM Eindeutigkeit im allgemeinen Fall gewährleistet. Alle anderen Nützlichkeitsmatrizen lassen hingegen verschiedene Gestaltungsmöglichkeiten zu.

2.5 Entscheidungstheorie und einige Gedanken über Ziel und Zweck der Strafe

Über Ziel und Zweck der Strafe sind, wie bereits erwähnt, verschiedene Auffassungen möglich. So kann man darunter den (materiellen oder ideellen) Ausgleich des Übels seitens des Täters im Rahmen des gesetzlich Zulässigen und Gebotenen verstehen, Vergeltung oder Verbrechensbekämpfung und -verhütung¹⁾ mittels (z.B.) General- oder Spezialpräventionen. Die strafrechtlichen Auffassungen bzgl. der Zielvorstellung, die auf die konkrete Gestalt einer Nützlichkeitsmatrix Einfluß ausüben können, haben wir auf S.308 in 4 Kategorien eingeteilt: (1) resozialisierende (wiedereingliedernde), (2) vergeltende (rächende), (3) (general- bzw. spezial-)präventive (abschreckende) und (4) wiederherstellende/wiedergutmachende (ausgleichende) Gerechtigkeits- Vorstellungen.

Der Unterschied zwischen der vergeltenden und der wiedergutmachenden Auffassung ist nicht so sehr in der Wirkung als vielmehr in der moralischen Dimension zu suchen. So rächt sich die Gesellschaft im ersten Fall an dem Einzelnen, wenn dieser sich ein Fehlverhalten erlaubt, während im zweiten Fall versucht wird, den Schaden des Opfers wiedergutzumachen um der sozialen Gerechtigkeit willen²⁾, die es erst möglich macht, daß die Interessen und Grundrechte³⁾ eines jeden, auch eines Machtlosen und weniger Privilegierten, geschützt sind.

Die Strafe als Ausgleich des Übels richtet sich einzig und allein an die Tatschuld, d.h. das begangene Unrecht (strafbarer oder objektiver Tatbestand) + Schuld des Angeklagten (subjektiver Tatbestand). Dagegen führen Verbrechensbekämpfung und -verhütung als Strafziele ihrerseits zu verschiedenen gesellschaftlichen Konsequenzen, die je nachdem, wie man sie interpretiert, sogar u.U. unerwünscht

1) Vgl. Haag, Rationale Strafzumessung (1970), S. 19.

2) Es ist allerdings eine offene Frage, ob die soziale Gerechtigkeit gewährleistet wird, indem der Täter den Schaden des Opfers wiedergutmacht, oder vielmehr, indem die Gesellschaft dem Opfer ein dem Unrecht formal äquivalentes materielles oder ideelles Gut zur Verfügung stellt. Würde man die soziale Gerechtigkeit im letzteren Sinn auffassen, so bedürfte es offenbar keines juristischen Bestrafungsapparates mehr.

3) Zu den Grundrechten gehört z.B. das Recht auf Leben, Freiheit, auf die individuelle Entfaltung und Abgrenzung etc.

sein können: Faßt man die Verbrechensbekämpfung allgemein auf, so wäre eine solche evtl. auch allein durch Allgemeinabschreckung erreichbar. Faßt man die Verbrechensverhütung eng (als Resozialisierung) auf, so wäre eine solche evtl. auch ohne Bestrafung des Täters erreichbar. In beiden Fällen würde man die Sachlage völlig verfehlen, nämlich, daß ein Unrecht in der Welt durch den Täter verwirklicht wurde und daß der Täter dieses Unrecht wiedergutmachen soll.

Verschiedene Auffassungen über Ziel und Zweck der Strafe bzw. die Methoden zu ihrer Realisierung werden im folgenden systematisch angegeben.

(1) Gerechtigkeit

Dieser Auffassung¹⁾ liegt das Prinzip des gerechten (angemessenen) Ausgleichs des Übels seitens des Täters (d.h. die Wiedergutmachung) im Rahmen des gesetzlich Gebotenen und Zulässigen zugrunde. Dabei darf die Strafe die Schwere der verschuldeten Straftat weder über- noch unterschreiten, d.h. die gewählte Maßnahme darf nicht härter sein als das begangene Unrecht, aber auch nicht milder. In der juristischen Literatur hat sich für diesen Zusammenhang ein bestimmter Ausdruck eingebürgert (von dem wir in dieser Arbeit bereits mehrmals Gebrauch machten), der der "tatschuldangemessenen Bestrafung". Alle tatschuldangemessenen Maßnahmen für entsprechende Tatbestände müssen gleichermaßen bewertet werden, d.h. die Bestrafung des Täters entsprechend dem verwirklichten Tatbestand muß ebenso wichtig sein wie die Nicht-Bestrafung des Nicht-Täters (z.B. muß die lebenslange Strafe für Mord ebenso wichtig sein wie etwa die Freiheitsstrafe von "nicht unter einem Jahr" für Raub oder der Freispruch für Notwehr). Allgemein sprechen wir hier davon, daß die Konsequenzen der tatschuldangemessenen Bestrafungspraxis wertneutral sind. Aus gesetzgeberischer Sicht ist diese Voraussetzung stets erfüllt²⁾. Die mildernden Umstände dürfen nur im Hinblick auf das Strafmaß bei der Konkretisierung des Strafrahmens, d.h. nur im Rahmen der gewählten Maßnahme berücksichtigt werden, jedoch nicht auf die Wahl der Maßnahme Einfluß haben³⁾. Die Nützlichkeitsmatrix, welche

1) Vgl. S.308, S.324, S.326f dieser Arbeit.

2) Vgl. S.323, Fn.1 dieser Arbeit.

3) Auf diesen Aspekt kommen wir im letzten Kapitel dieses Teils noch zurück.

diese Auffassung repräsentiert, sei hier noch einmal angeführt¹⁾

$$\text{GM:} \quad \begin{matrix} & T_1 & T_2 & \dots & T_n \\ \begin{matrix} M_1 \\ M_2 \\ \vdots \\ \vdots \\ M_n \end{matrix} & \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} \end{matrix}$$

Die obige Matrix garantiert auch in zwei anderen Punkten die formale Gerechtigkeit: Sie ist eindeutig; so folgt aus ihr die (juristisch geforderte) Gleichbehandlung, denn auf individuelle Unterschiede wird hier keine Rücksicht genommen und damit der rechtliche Grundsatz der Gleichbehandlung vor dem Gesetz nicht verletzt; zweitens wird hier eine Maßnahme als die beste abgeleitet, der nicht ein wenig wahrscheinlicher Tatbestand zugrunde liegt.

Man kann sich fragen, ob die hier dargestellte Gerechtigkeitsauffassung in der Praxis eine reale Entsprechung besitzt, oder ob sie doch eher einen normativen Charakter trägt, denn in der Rechtspraxis spricht man von Gerechtigkeit eher in Verbindung mit der Gleichbehandlung und nicht als strafrechtlicher Zielsetzung an sich. Ich will hier allerdings die Behauptung aufstellen, daß der oben dargestellten Gerechtigkeitsauffassung etwas Fundamentales anhaftet, das im Wesen der Bestrafungspraxis selbst verankert ist und mit der sozialen Organisation der Gesellschaft zu tun hat. Die Gesellschaft erwartet ein Wohlverhalten eines jeden gegenüber dem anderen bzw. der Gruppe an sich, um das gewisse soziale Gleichgewicht, das ein Zusammenleben in der Gesellschaft erst möglich macht, aufrechtzuerhalten. Zerstört jemand durch sein Fehlverhalten den von der Gesellschaft erwünschten sozialen Zustand, so muß er in einer dem Fehlverhalten angemessenen materiellen oder ideellen Weise und in einer von der Gesellschaft zu diesem Zweck errichteten institutionellen Form (dem Justizapparat) sein Verfehlen ausgleichen, d.h. wiedergutmachen, und dadurch das soziale Gleichgewicht wiederherstellen. Der höchste juristische Nutzen entspringt demnach daraus,

1) Vgl. S. 326; wir beschränken uns hier wiederum auf eine $n \times n$ -Matrix, d.h. es kommt jedem Tatbestand in dieser Matrix nur eine einzige Maßnahme zu.

daß auf das Fehlverhalten in einer juristisch gebotenen und zulässigen (angemessenen) Weise geantwortet wird, wobei auf individuelle Unterschiede keine Rücksicht genommen wird, sondern einzig und allein auf die Schwere des Vergehens (auf den objektiven Sachverhalt).. Eine dem Fehlverhalten nicht angemessene Behandlung (sei sie härter oder milder) wird als ungerecht empfunden. Ob die konkreten (gesetzlich angegebenen) Strafen angemessen sind und das soziale Gleichgewicht durch die Bestrafungspraxis tatsächlich wiederhergestellt wird, bzw. ob die Bestrafungspraxis an sich überhaupt gerecht ist, sind keine entscheidungstheoretischen, sondern zum Teil ethische, zum Teil empirische Fragen, auf die hier selbstverständlich keine Antwort gegeben werden kann. Das ist auch ein weiterer Grund dafür, warum wir die juristische Gerechtigkeit im Sinne der formalen Wiederherstellung des juristischen Status Quo verstanden haben¹⁾.

So gesehen können wir auch nicht von "mehr oder weniger gerecht" sprechen, indem wir etwa eine mildere Bestrafung, als es dem Tatbestand entspricht, für gerechter halten als eine härtere für den gleichen Tatbestand, denn durch eine mildere bzw. härtere Behandlung wird das soziale Gleichgewicht bzw. der juristische Status Quo nicht wiederhergestellt.

Da der Gerechtigkeitsauffassung etwas Fundamentales anhaftet, was mit einer konkreten Entscheidungssituation wenig zu tun hat, ist sie auch in einer gewissen Hinsicht mangelhaft: So werden von ihr bestimmte individualisierende und für die moderne Rechtssprechung wichtige (wie z.B. resozialisierende) Gedanken nicht berücksichtigt.

(2) Vergeltung

Dieser Auffassung liegt das Vergeltungsprinzip zugrunde, wonach die Strafe weder das Maß der Tatschuld überschreiten (das sog. "Übermaßverbot") noch unterschreiten darf²⁾. In dieser Hinsicht ähnelt das Vergeltungsprinzip dem Prinzip des gerechten Ausgleichs des Übels. Zum unterschiedlichen moralischen Hintergrund kommt allerdings noch ein weiterer Aspekt hinzu: Das Vergeltungsprinzip

1) Der erste Grund besteht (vgl. S. 308, Fn. 2) darin, daß nicht tatsächlich der vorunrechtliche Zustand (Grundzustand) wiederhergestellt wird (dies ist nur in den seltensten Fällen möglich), sondern ein an sich neues, jedoch dem alten formal äquivalentes, soziales Gleichgewicht errichtet wird.

2) Vgl. Haag, S. 19.

verlangt offensichtlich nicht, daß alle tatschuldangemessenen Behandlungen gleichermaßen bewertet werden, d.h. daß sie wertneutral sind. So kann man sich leicht vorstellen: je schwerer das Verfehlen ist, um so wünschenswerter erscheint seine Bestrafung bzw. seine Sühne, etwa dann, wenn man die Bestrafung eines schweren Raubs für wichtiger hält als die eines geringen Diebstahls. Es kann daher passieren, daß das Gewicht zugunsten einer bestimmten Maßnahme für den entsprechenden Tatbestand verschoben wird, wie etwa bei der folgenden Matrix:

$$\begin{array}{c}
 \begin{array}{c} M_1 \\ M_2 \\ \cdot \\ \cdot \\ \cdot \\ M_n \end{array}
 \begin{array}{c} T_1 \quad T_2 \quad \dots \quad T_n \\ \left[\begin{array}{cccc} 1 & 0 & \dots & 0 \\ 0 & x & \dots & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \dots & y \end{array} \right] \end{array}
 \end{array}
 \quad \text{mit } x, y \neq 0,1$$

Aus dieser Matrix ersieht man, daß zwar das Übermaßverbot gilt, doch haben die tatschuldangemessenen Maßnahmen für die ihnen entsprechenden Tatbestände nicht alle den gleichen Wert (wie bei der GM). So hat die härteste Maßnahme für den schwersten Tatbestand den höchsten Nützlichkeitswert, dagegen die mildere (und sogar die mildeste) Maßnahme für den leichteren Tatbestand einen Nützlichkeitswert ungleich 1. Nach dieser Matrix kann eine härtere Maßnahmen u.U. auch dann gewählt werden, wenn der dieser Maßnahme entsprechende Tatbestand nicht mit überwiegender Wahrscheinlichkeit festgestellt ist. Darin äußert sich im Grunde genommen der Vergeltungscharakter dieser Auffassung. Als Extremfälle der Vergeltung können somit die Gerechtigkeit (wenn alle Nützlichkeitswerte auf der Hauptdiagonale von links nach rechts gleich sind) und die Ungerechtigkeit (wenn ein noch so minimaler Wahrscheinlichkeitswert für den Tatbestand ausreicht, um die diesem Tatbestand entsprechende Maßnahme zu wählen) gelten¹⁾.

Die obige Matrix garantiert außerdem nicht die Gleichbehandlung (sie ist nicht eindeutig), denn man kann abhängig von Fall und Individuum bestimmen, welche Bestrafung der Vergeltung am besten entspricht.

1) Das letztere wird übrigens durch die Annahme des noch zu formulierenden Postulats JP2 (vgl. S.352) verhindert. Die Vergeltungsauffassung ist dadurch sozusagen "entschärft"; man kann zwar auch weiterhin von der Vergeltung bei der Strafbestimmung ausgehen; die ungerechten Konsequenzen können aber nicht mehr abgeleitet werden.

In unserer Sprechweise können wir somit sagen, daß die Vergeltungskonzeption zwar wie die Gerechtigkeitsauffassung eine härtere bzw. eine mildere als die tatschuldangemessene Behandlung verbietet, die Konsequenzen der tatschuldangemessenen Bestrafung jedoch nicht alle als wertneutral postuliert.

(3) Prävention

1. Die generalpräventive Auffassung

Dieser Auffassung liegt das Prinzip der Allgemeinabschreckung zugrunde. Dies kann man aber auch dann erreichen (wenn überhaupt), wenn das Strafmaß überschritten wird (graphisch: positive Nützlichkeitswerte oberhalb der Hauptdiagonale von links nach rechts). Die generalpräventive Auffassung verletzt in bezug auf die Wahl der Maßnahme den in der Rechtspraxis verankerten Grundsatz, wonach die Strafe aus generalpräventiven Gründen nicht über das Maß der Tatschuld hinaus erhöht werden darf¹⁾. Generalpräventive Erwägungen dürfen jedoch nach diesem Grundsatz innerhalb der gewählten Maßnahme bei der Konkretisierung des Strafrahmens Berücksichtigung finden²⁾. Wir werden später zeigen³⁾, daß sie auch hier unzulässig sind, sofern dadurch die tatschuldangemessene Strafe erhöht wird. Beachtet man dagegen diesen Grundsatz bei der Wahl der Maßnahme, so wird der Unterschied zwischen der Vergeltungs- und der generalpräventiven Auffassung hinfällig⁴⁾.

2. Die spezialpräventive Auffassung (Einzelabschreckung)

Diese Auffassung geht vom Prinzip der Einzelabschreckung aus. Das kann allerdings, wie bei der generalpräventiven Auffassung, dadurch erreicht werden, daß das Übermaßverbot verletzt wird, wie in der folgenden Situation:

Ein Richter ist der Ansicht, daß man einen reichen Parksünder von

1) Dieser Grundsatz (das "Übermaßverbot") besagt folgendes: Wird der Angeklagte für schuldig befunden (und das setzt natürlich voraus, daß seine Täterschaft nachgewiesen ist), so muß die Maßnahme gewählt werden, die dem objektiven Tatbestand entspricht, d.h. sie darf nicht härter ausfallen als die für einen solchen Tatbestand durch das StGB vorgesehene Maßnahme. Kurz und prägnant wird dieser Gedanke durch folgenden Satz ausgedrückt: "Keine Strafe ohne (entsprechendes) Verbrechen".

2) Vgl. Haag, S. 20ff.

3) Vgl. S. 405 dieser Arbeit.

4) Später wird gezeigt, daß die generalpräventive Auffassung nach JP2 ebensowenig möglich ist wie die ungerechte Vergeltung.

seinen 'Untaten' abschrecken kann, wenn dieser mit einer für seine Verhältnisse 'empfindlichen' Geldstrafe von 1000 DM bestraft wird. Dieser Richter verletzt dann das Prinzip des Übermaßverbots, indem er eine Strafe wählt, die bei weitem das verwirklichte Unrecht bzw. den Schaden, der durch diese Tat der Gesellschaft entstanden ist, übersteigt.

Spezialpräventive Erwägungen (im Sinn der Einzelabschreckung) dürfen somit ebenso wenig wie generalpräventive Gesichtspunkte die Wahl der Maßnahme (des Strafrahmens) beeinflussen. Sie können jedoch bei der Bestimmung der Strafhöhe innerhalb der gewählten Maßnahme eine Rolle spielen¹⁾. Dadurch wird dem Umstand Rechnung getragen, daß bestimmte Eigenschaften des Täters, wie etwa, daß es sich um einen Wiederholungstäter handelt, bei der Bemessung der Strafhöhe gewöhnlich eine Rolle spielen.

Sowohl die general- wie auch die spezialpräventive Auffassung sind also sowohl in juristischer wie auch in formaler Hinsicht ungerecht: Sie garantieren weder die Gleichbehandlung vor dem Gesetz noch, daß die beste Maßnahme nicht nach einem unwahrscheinlichen Tatbestand abgeleitet werden soll. Darüber hinaus sind sie deshalb grundsätzlich ungerecht, weil sie ihrerseits das soziale Gleichgewicht bzw. den juristischen Status Quo verletzen, indem sie von sich aus eine ungerechte (und im Rahmen der präventiven Auffassung offensichtlich nicht mehr wiedergutzumachende) Handlung/Zustand (eine härtere Bestrafung oder die Bestrafung eines Unschuldigen) akzeptieren.

(4) Resozialisation

Dieser Auffassung liegt das Prinzip der Wiedereingliederung des Täters in die Gesellschaft²⁾ zugrunde. Das kann allerdings auch dann erreicht werden, wenn der Täter milder als es dem Tatbestand entspricht oder überhaupt nicht bestraft wird. Eine solche Auffassung kann beispielsweise durch die folgende Nützlichkeitsmatrix dargestellt werden:

1) insofern auch hier nicht die Höhe der tatschuldangemessenen Strafe überschritten wird.

2) Gemeint ist hier: die Wiedereingliederung in die Gesellschaft (oder "Besserung" bzw. "soziale Heilung") des Täters mit 'humanen' Mitteln.

ten individuellen Gründen) milder bestraft, als es der Tatschuld entspricht.

2.6. Die Hierarchie juristischer Ziele

Insgesamt läßt sich eine Hierarchie juristischer Ziele aufbauen. Diese Ziele können dabei in einer doppelten Funktion fungieren: zum einen als eigentlicher Zweck der juristischen Behandlung und zum anderen als das Mittel (die Art und Weise), diesen Zweck zu erreichen. Im folgenden wird eine Übersicht über die Ziele in dieser doppelten Funktion gegeben.

Das fundamentale Ziel ist die Gerechtigkeit, d.h. die formale Wiederherstellung des Grundzustands des sozialen Gleichgewichts bzw. des juristischen Status Quo (1).

Zu den primären Zielen gehören: Verbrechensbekämpfung und -verhütung (2), Vergeltung (3) und Genugtuung des Geschädigten (4).

Zu den sekundären Zielen gehören: Resozialisation (5), Abschreckung (6), die Befriedigung des Sicherheitsbedürfnisses der Gesellschaft (7) und die Leidensverringerung des Bestraften (8).

Die Mittel zur Erreichung der obigen Ziele kann man in 4 verschiedene Kategorien unterteilen: (a) Wiedergutmachung/Ausgleich¹⁾, (b) Resozialisation, (c) Vergeltung und (d) Prävention (Einzel- bzw. Allgemeinabschreckung).

Die genaue Zuordnung zwischen Zielen und möglichen Mitteln zeigt die folgende systematische Übersicht:

¹⁾ Gemeint ist hier: der gerechte Ausgleich, die gerechte Behandlung bzw. juristisch ausgedrückt: die tatschuldangemessene Behandlung.

- (1) durch (a);
- (2) durch (a), (b), (c), (d);
- (3) durch (c);
- (4) durch (a), (c), (d);
- (5) durch (b), aber auch (a);
- (6) durch (c), (d);
- (7) durch (a), (b), (c), (d);
- (8) durch (b), allerdings ist nicht auszuschließen, daß auch (a) - die gerechte Bestrafung - eine Art Sühnewirkung (als moralische Leidensverringering) auf manchen Angeklagten haben kann.

Auf ein weiteres für das Prozeßverfahren wichtiges Ziel, die Kostenminimierung, wird hier nicht eingegangen, denn dieses Ziel betrifft nicht die Bewertung der einzelnen Konsequenzen, sondern die des gesamten Entscheidungsprozesses. Die Frage muß hier lauten: Soll man die Suche nach weiteren Informationen (Beweismitteln) abbrechen (weil ihre Beschaffung kostspielig ist) oder soll man weiter forschen (weil andere Gesichtspunkte wichtiger sind). Allerdings kann die Kostenfrage auch im Zusammenhang mit der Entscheidung auftreten, ob man den Angeklagten zu einer Gefängnisstrafe verurteilen soll (was dem Staat Unterbringungskosten verursacht) oder 'nur' zu einer Geldstrafe (wodurch der Staat profitiert) oder ihn freisprechen. Es ist klar, daß ein solches Ziel nicht primär sein kann. Es ist sogar fraglich, ob man es zu den sekundären Zielen rechnen kann: Indem der Staat den Angeklagten freispricht, nur weil ihm die Unterbringung Geld kosten würde bzw. weil er Geld sparen will, verfolgt er ein höchst inhumanes und ungerechtes Ziel. Allerdings können wir (vorausgesetzt, daß es sich um geringe Vergehen handelt) innerhalb der Resozialisationsauffassung auch diese Zielsetzung berücksichtigen (etwa in einer Nützlichkeitsmatrix, indem dem Freispruch bzw. den milderen Strafen (wie z.B. eine Geldstrafe) positive Nützlichkeitswerte zugeordnet werden).

Zusammenfassung und Ausblick

In diesem Paragraphen wurde der Versuch unternommen, den juristischen Nutzen durch die Theorie eines solchen einzugrenzen. Folgendes wurde festgestellt: es darf sich bei diesem nicht um einen persönlichen Nutzen des Individuums (des Richters) handeln, sondern nur um einen sozialen Nutzen, der die Belange der von der Straftat Betroffenen (v.a. des Opfers und seiner Angehörigen) und der Gesellschaft berücksichtigt. Die Nützlichkeitszuordnung ist aber auch dann subjektiv: es ist ein Individuum (der Richter), das beurteilt, welche Konsequenzen seiner Handlungen von mehr oder weniger großem sozialen Nutzen sind. Damit die Interpretation des juristischen Nutzens - auch wenn er sozial verstanden wird - nicht zu mannigfaltig ausfällt und um außerdem der Theorie des juristischen Nutzens eine einigermaßen gesicherte Grundlage zu bieten, haben wir an die in der Rechtspraxis vorhandenen strafrechtlichen Theorien über den Sinn und Zweck der Strafe angeknüpft. Es stellte sich jedoch heraus, daß der von uns als Gerechtigkeitsauffassung bezeichnete Grundgedanke, der davon ausgeht, daß eine Straftat eine (negative) Veränderung des sozialen Gleichgewichts oder des juristischen Status Quo darstellt und die Strafe als die formale Wiederherstellung dieses Sta-

tus Quo begreift, in der Rechtsliteratur noch nicht in dieser Form formuliert wurde. Von diesem fundamentalen Ziel ausgehend - welches auch als einziges die Eindeutigkeit der Nutzenzuordnung und damit den juristischen Grundsatz der Gleichbehandlung vor dem Gesetz garantiert - haben wir andere juristisch geläufige Ziele in ihrer doppelten Funktion: zum einen als Zweck der Strafe und zum anderen als Mittel zu dessen Erreichung, formal untersucht.

Unsere Untersuchung strafrechtlicher Ziele war auch aus folgendem Grunde rein formal: Wir haben uns keine Gedanken darüber gemacht, warum auf einen bestimmten Tatbestand diese oder jene durch das StGB festgelegte Strafe (Rechtsfolge) folgt, warum z.B. auf Mord "lebenslänglich" steht oder warum der Schwangerschaftsabbruch gemäß §218 StGB bestraft wird oder überhaupt, warum das Bestrafen gut oder schlecht ist. Auch haben wir bestimmte rechtliche Grundsätze, wie Übermaßverbot und Gleichbehandlung vor Gericht so, wie sie in der Rechtspraxis sind, übernommen, ohne sie extra in Postulaten zu formulieren (es zeigt sich später, daß das Übermaßverbot ohnehin aus dem Postulat JP2 folgt; die Gleichbehandlung wird stets von der Rechtssprechung selbst verletzt, etwa durch die Resozialisierungstheorie) und ohne auf ihre ethische Rechtfertigung einzugehen. So gesehen entbehren auch die in dieser Arbeit dargestellten Auffassungen - wie Gerechtigkeits-, Vergeltungs-, Resozialisations- und Abschreckungsauffassung - einer ethischen Grundlage. Um sich mit solchen Fragen sinnvoll beschäftigen zu können, muß uns eine rationale juristische Ethik zur Verfügung stehen, die sich mit Normen und Wertungen vom rationalen Standpunkt aus auseinandersetzt. Ohne eine solche Ethik ist es unmöglich, die rechtlichen Gedanken auf ihre ethische Relevanz hin zu überprüfen.

Wie in diesem Paragraphen auch gezeigt wurde, können fast alle Ziel-auffassungen (außer der Gerechtigkeitsauffassung) ungerechte Ergebnisse liefern, wenn man sie nicht in einer bestimmten Weise einschränkt; in allen diesen Auffassungen kann eine Maßnahme als die beste abgeleitet werden, die gar nicht dem wahrscheinlichsten, gelegentlich sogar dem unwahrscheinlichsten Sachverhalt entspricht. Im nächsten Paragraphen wird u.a. auch deswegen eine Forderung (in Form einer Voraussetzung an die entscheidungstheoretische Beurteilung) an den Wahrscheinlichkeitswert des (strafbaren) Tatbestands aufgestellt, die die Ableitung derartig ungerechter Ergebnisse weitgehend einschränkt.

Eine andere Möglichkeit, nämlich stattdessen die Nützlichkeitsmatrix einzuschränken, erscheint aus drei Gründen weniger erfolgversprechend: erstens wird durch die Einschränkung der Nützlichkeitsmatrix auch die Vielfalt an juristischen Auffassungen bzgl. des Ziels und Zwecks der Strafe eingeschränkt und sogar ihrer Bedeutung durch eine Matrix entzogen, deren normative Implikationen weniger akzeptabel bzw. einsichtig sind; zweitens wird dem für die Rechtspraxis wichtigen und plausiblen Gedanken, daß der Sachverhalt für sich (d.h. ohne daß vorerst die Entscheidungsfrage erläutert wird) mit einer bestimmten (genügend hohen) Wahrscheinlichkeit nachgewiesen sein soll, nicht mehr Rechnung getragen¹⁾ und drittens ist anzunehmen, daß diese Einschränkung sich für größere als 2x2-Nützlichkeitsmatrizen überhaupt nicht angeben läßt.

1) Durch die Vermengung der wahrscheinlichkeitstheoretischen mit der entscheidungstheoretischen Beurteilung steht dann nicht der Nachweis des Sachverhalts, sondern die Ableitung einer Entscheidung im Mittelpunkt der strafrechtlichen Tätigkeit.

§3. Axiomatische Festlegung einer rationalen juristischen Entscheidungstheorie

Dieser Paragraph ist insofern zentral für die ganze Arbeit, als hier zusätzlich zu den entscheidungstheoretischen Postulaten einige Forderungen aufgestellt und begründet werden, die aus dem entscheidungstheoretischen Grundmodell (Savage-Modell) eine juristische Entscheidungstheorie entstehen lassen. Wie im ersten Kapitel dieses Teils gezeigt wurde, ist das Grundmodell für die juristischen Zwecke zu allgemein.

Insgesamt werden vier Postulate (eines davon in der Konditionalform) formuliert, die die rationale Entscheidungsbildung juristisch einschränken sollen. Bei diesen vier Postulaten handelt es sich um minimale Forderungen, die zwar bestimmte, aus dem Grundmodell sich ergebende nicht befriedigende Folgen verhindern können und auch den subjektiven Sachverhalt in die rationale Entscheidungsbildung einbeziehen, gewisse juristische Forderungen aber, etwa die, daß die Entscheidung nur dann ableitbar sein soll, wenn der entsprechende Sachverhalt mit "an Sicherheit grenzender Wahrscheinlichkeit" festgestellt ist, nicht erfüllen. Es wird außerdem gezeigt, daß die letztere Forderung auch formal zu unsinnigen Einschränkungen bzgl. der Nützlichkeitsmatrix führt.

Von den vier vorgeschlagenen Postulaten erscheint das erste Postulat JP1, das die Indifferenzbedingung in bezug auf die juristischen Maßnahmen aufhebt, am plausibelsten und wird wohl kaum auf Ablehnung seitens der Juristen stoßen.

Das zweite Postulat JP2 (der Grundsatz der überwiegenden kumulativen Wahrscheinlichkeit) besagt, daß der strafbare Tatbestand, der der juristischen Entscheidungsbildung zugrunde gelegt wird, mit kumulativ überwiegender Wahrscheinlichkeit festgestellt werden soll. Es besagt aber nicht, mit welcher Wahrscheinlichkeit dieser festgestellt werden soll (daher wird diese Forderung als minimale Forderung an den strafbaren Sachverhalt bezeichnet; d.h. die "kumulativ überwiegende Wahrscheinlichkeit" bildet gewissermaßen eine "untere Schranke" für die entscheidungstheoretische Beurteilung und entsprechend die obige Forderung JP2 die Voraussetzung für die Anwendung der Entscheidungstheorie auf die Strafrechtspraxis).

Das dritte Postulat JP3 (das sog. Gerechtigkeitspostulat) muß nur dann akzeptiert werden, wenn man bei der Entscheidungsbildung von der Fundamentalkonzeption (Gerechtigkeitsauffassung ausgeht); daher wird dieses Postulat in der Konditionalform formuliert (Dieses Postulat ist lediglich von theoretischem Interesse, da in der Praxis kaum zu erwarten ist, daß der Entscheidung die Gerechtigkeitsmatrix (S.331 dieser Arbeit), die die Gleichbehandlung auf Kosten der Individualisation garantiert, in ihrer reinen Form zugrunde liegen wird).

Das Postulat JP4 schließlich bezieht in die rationale juristische Entscheidungsbildung den subjektiven Sachverhalt (die Schuld- bzw. Schuldfähigkeit des Angeklagten) mit ein. Da die Entscheidung nach diesem Postulat erst dann eintritt, wenn bereits die Wahl der dem objektiven Tatbestand entsprechenden Maßnahme nach JP1-JP3 erfolgt

ist, hat dieses Postulat lediglich eine die Entscheidung bestätigende Funktion.

Durch JP1-JP4 wird somit der ganze juristische Entscheidungsprozeß - von der Feststellung des objektiven Tatbestands bis zur Wahl der Maßnahme und der Bestätigung dieser Wahl nach dem subjektiven Tatbestand - axiomatisch festgelegt.

Die Postulate werden für beliebig endlich viele Tatbestände und beliebig endlich viele Maßnahmen formuliert.

Bevor wir zur Formulierung der vier Postulate übergehen, wollen wir hier noch einige wichtige Voraussetzungen für die Art der Darstellung explizit angeben:

1. Sowohl die Tatbestände wie auch die Maßnahmen werden graphisch geordnet, indem die Tatbestände von links nach rechts und die Maßnahmen von oben nach unten nach abnehmender Schwere¹⁾ geordnet angeführt werden (so wiegt "Mord" schwerer als "Totschlag", deswegen wird er links von diesem angeführt; entsprechendes gilt für "lebenslänglich" und "Freiheitsstrafe 'bis zu fünf Jahren'": die erste Maßnahme wird oberhalb der zweiten angeführt). Von einer solchen graphischen Ordnung haben wir bereits bei unserer bisherigen Darstellung implizit Gebrauch gemacht. Sie erleichtert nicht nur die Formulierung der Axiome, sondern auch die Handhabung einer juristischen Entscheidungssituation im Modell.
2. Bei den $T_1, \dots, T_n$ Tatbeständen, für die die Postulate formuliert werden, handelt es sich nicht um alle überhaupt möglichen Tatbestände, sondern nur um solche, die in einer konkreten juristischen Entscheidungssituation (in einem konkreten Rechtsfall) für die juristische Beurteilung in Frage kommen.
3. Bei den $T_1, \dots, T_n$ in den Postulaten JP1-JP3 handelt es sich einzig und allein um objektive Tatbestände (wie sie etwa im StGB formuliert sind).
4. Bei den $M_1, \dots, M_r$ in JP1-JP3 handelt es sich ausschließlich um Strafrahen; es wird noch nichts über die konkrete Strafdauer oder über die konkrete Geldstrafe ausgesagt (zu diesem Punkt vgl. 5 Kap.).
5. $P(T_1), \dots, P(T_n)$ sind im folgenden Abkürzungen für bedingte Wahrscheinlichkeiten.

1) "Schwere" in Bezug auf die Tatbestände ist hier also im Sinne des "Unrechtsgehalts" zu verstehen.

3.1 Aufhebung der entscheidungstheoretischen Indifferenzbedingung

Kehren wir zu unserem Beispiel auf S.300ff zurück. Dort waren 5 Tatbestände möglich: "schwerer Raub" (T_1), "Raub" (T_2), "Diebstahl mit Waffen" (T_3), "Diebstahl" (T_4) und "kein Diebstahl" (T_5) und entsprechend 5 Maßnahmen: "nach §250 (wegen schweren Raubs)" (M_1), "nach §249 (wegen Raubs)" (M_2), "nach §244 (wegen Diebstahls mit Waffen)" (M_3), "nach §242 (wegen Diebstahl)" (M_4) und "Freispruch" (M_5). Nach der Berechnung erhielten die beiden Maßnahmen M_1 ("Bestrafung wegen schweren Raubs") und M_3 ("Bestrafung wegen Diebstahls mit Waffen") den gleichen höchsten Erwartungswert des Nutzens. Aus entscheidungstheoretischer Sicht könnte man in einem solchen Fall jede dieser beiden Maßnahmen wählen. Das fanden wir aber wenig befriedigend (vgl.S.305), denn juristisch gesehen macht es einen Unterschied, ob man jemanden wegen "schweren Raubs" oder wegen "Diebstahls mit Waffen" verurteilt. Noch fataler wäre jedoch die Situation, wenn etwa sowohl M_1 ("Bestrafung wegen schweren Raubs") als auch M_5 ("Freispruch") den höchsten Erwartungswert des Nutzens erhielten.

Um einer solchen Situation juristisch gerecht zu werden, stellen wir hier eine Forderung auf, nach der (aus humanitären Gründen) einer minder schweren Maßnahme der Vorzug gegeben wird, wenn eine entscheidungstheoretische Indifferenz vorliegt. Diese juristische Forderung sei im folgenden Postulat explizit formuliert¹⁾:

JP1 Seien $M_1, \dots, M_r$ Maßnahmen, die nach ihrer Schwere geordnet sind, d.h. M_1 ist die schwerste Maßnahme und M_r die leichteste, $E_u(M_1), \dots, E_u(M_r)$ seien die entsprechenden Erwartungswerte des Nutzens. Dann gilt:

Falls $E_u(M_i) = E_u(M_j)$, dann $M_j > M_i$, $i, j = 1, \dots, r; i < j$.

1) Hier und im folgenden werden wir von r (mit $r \geq n$) anstatt von n Maßnahmen, die den n Tatbeständen zugeordnet sind, sprechen. Dadurch soll (wie bereits auf S.289 erläutert wurde) der Tatsache Ausdruck verliehen werden, daß man manche Maßnahmen wiederum in verschiedene Teilmaßnahmen aufspalten kann (wie etwa in Gefängnisstrafe, Geldstrafe, Bewährungsstrafe und dergl.); z.B. läßt sich im obigen Beispiel die Maßnahme "nach §242 (wegen Diebstahls)" (M_4) in "Freiheitsstrafe bis zu fünf Jahren" (M_4^1) und "Geldstrafe" (M_4^2) aufspalten.

Schauen wir uns unser Beispiel noch einmal an. Unter Beachtung von JPl werden nicht mehr die beiden Maßnahmen M_1 und M_3 als die besten abgeleitet, sondern einzig und allein die "mildeste" von beiden M_3 , d.h. "Bestrafung wegen Diebstahls mit Waffen".

3.2. Beweisgrenzen. Die Rechtfertigung des Beweiskriteriums "mit überwiegender Wahrscheinlichkeit" für die 2x2-Konsequenzenmatrix

Im Beispiel auf S.303ff haben wir "Bestrafung wegen Mordes" als die beste Maßnahme abgeleitet, obwohl der entsprechende Tatbestand "Mord" die geringste Wahrscheinlichkeit besaß, d.h. alle anderen Tatbestände: "Totschlag" (T_2), "minder schwerer Fall des Totschlags" (T_3) und "keine Tötung" (T_4) waren nicht weniger wahrscheinlich als T_1 (T_3 war gleichwahrscheinlich, T_2 und T_4 wahrscheinlicher als T_1). Im folgenden wird eine Forderung aufgestellt, die zur Einschränkung derartig unbefriedigender Ergebnisse führt.

Zunächst wollen wir aber untersuchen, wie die numerische Darstellung verschiedener Beweiskriterien für den strafbaren Tatbestand¹⁾ sich auf die Gestalt der Nützlichkeitsmatrix auswirkt. Wir können nämlich, falls uns die Wahrscheinlichkeit für den Sachverhalt bekannt ist, mit einer weiteren Voraussetzung (= der Annahme der Indifferenz bzgl. der Maßnahmen) die Nützlichkeitsmatrix, die einer solchen Entscheidung zugrunde liegt, im einfachsten 2x2-Matrizenfall ableiten (falls noch einige weitere Bedingungen erfüllt sind).

Wir beginnen mit dem in der Jurisprudenz geltenden Grundsatz, daß der Beweis für den strafbaren Sachverhalt zumindest mit einer "an Sicherheit grenzenden Wahrscheinlichkeit" geführt werden muß. Drü-

1) wie etwa der Grundsatz "mit an Sicherheit grenzender Wahrscheinlichkeit" durch die reelle Zahl 0.99; aber auch andere hypothetisch angenommene Beweisgrenzen für die Wahrscheinlichkeit des strafbaren Sachverhalts sind möglich.

cken wir diesen Beweiswert etwa durch die reelle Zahl 0.99 aus¹⁾ so besagt die obige Forderung, daß die Entscheidung (Verurteilung des Angeklagten) entsprechend dem strafbaren Tatbestand nur dann abgeleitet werden soll, wenn mit einer Wahrscheinlichkeit von ≥ 0.99 festgestellt ist, daß der entsprechende Sachverhalt (d.h. daß der Angeklagte der Täter ist) gilt und andernfalls (d.h. für eine Wahrscheinlichkeit ≤ 0.99) eine andere Entscheidung (z.B. "Freispruch") abzuleiten ist²⁾

Wir haben im ersten Teil gezeigt, daß diese Forderung sowohl aus theoretischen Erwägungen wie auch aus praktischen Gründen unerfüllt bleiben muß. Jetzt zeigen wir, daß sie auch im Rahmen eines Entscheidungsmodells zu unerträglichen und unplausiblen Einschränkungen in bezug auf die Nützlichkeitszuordnung zu den Konsequenzen führt.

$p = 0.99$ (d.h. die Beweisgrenze liegt bei 0.99)

Im folgenden gehen wir davon aus, daß zwei Tatbestände T_1 und T_2 (z.B. "Diebstahl"/"kein Diebstahl") und entsprechend zwei Maßnahmen M_1 und M_2 ("Bestrafung wegen Diebstahls"/"Freispruch") in Frage kommen. Wir setzen hier noch einmal explizit voraus (im folgenden aber nicht mehr), daß die Konsequenzenmatrix durch die graphische Ordnung (vgl.S.341) festgelegt ist. Seien "x", "y", "z" und "w" Variablen für die jeweiligen Nützlichkeiten in der durch diese Konsequenzenmatrix bestimmten Nützlichkeitsmatrix, wobei diese wie folgt aussehe:

1) Wir analysieren hier also die in der Rechtsliteratur gegebene Interpretation des Beweiskriteriums "mit an Sicherheit grenzender Wahrscheinlichkeit", nämlich die als einem numerischen Wahrscheinlichkeitswert von 0.99 (vgl.S.58, Fn.1 dieser Arbeit; zu den zwei verschiedenen Interpretationen dieses Beweiskriteriums in der traditionellen Rechtsprechung siehe ebenda). Was die Interpretation des Begriffs "mit an Sicherheit grenzender Wahrscheinlichkeit" als einem kontexteliminierbaren Begriff betrifft, so erübrigt hier seine Analyse, denn, wie bereits gesagt (vgl.S.58, Fn.1) handelt es sich bei diesem um keinen Wahrscheinlichkeitsbegriff mehr.

2) Wir nehmen also an, daß evtl. sogar dann, wenn die Wahrscheinlichkeit für den strafbaren Tatbestand = 0.99 ist, eine mildere Maßnahme gewählt werden kann, dagegen nicht für eine Wahrscheinlichkeit > 0.99 , d.h. für $p = 0.99$ gilt die Indifferenz bzgl. der beiden Maßnahmen. Würde man darauf bestehen, daß auch für $p = 0.99$ nur die Verur-

$$\begin{bmatrix} x & y \\ z & w \end{bmatrix}$$

Um der Forderung "mit an Sicherheit grenzender Wahrscheinlichkeit" zu genügen, müssen offensichtlich die beiden folgenden Bedingungen gelten:

$$M_1 \succ M_2 \leftrightarrow 0.99 \cdot x + 0.01 \cdot y \geq 0.99 \cdot z + 0.01 \cdot w$$

$$M_2 \succ M_1 \leftrightarrow 0.99 \cdot z + 0.01 \cdot w \geq 0.99 \cdot x + 0.01 \cdot y$$

Nehmen wir (was plausibel erscheint) an, daß die Bestrafung des Nicht-Täters den Nutzen Null (d.h. $y = 0$)¹⁾ und die Nicht-Bestrafung des Nicht-Täters den Nutzen Eins (d.h. $w = 1$)²⁾ hat, dann reduziert sich das obige System auf die beiden Ungleichungen:

$$0.99 \cdot x \geq 0.99 \cdot z + 0.01$$

$$0.99 \cdot x \leq 0.99 \cdot z + 0.01$$

Ordnen wir x willkürlich den Wert 1 zu, so muß der Wert von z ungefähr 0.99 sein, um den beiden Bedingungen zu genügen. Die Nützlichkeitsmatrix hat also die Gestalt:

$$\begin{bmatrix} 1 & 0 \\ 0.99 & 1 \end{bmatrix}$$

D.h. das Entscheidungssubjekt muß die beiden Konsequenzen: "Strafe für die Straftat" und "keine Strafe für die Straftat" ungefähr gleich bewerten³⁾. Dies ist jedoch eine zu starke und unrealistische Bedingung, die sicher von vielen nicht akzeptiert wird.

teilung ableitbar sein soll, so müßte die Indifferenzbedingung bei einem Wahrscheinlichkeitswert < 0.99 gelten, d.h. das Individuum könnte bei diesem Wert (< 0.99) jede der beiden Maßnahmen als die beste wählen. Aus diesem Grund scheint die obige Interpretation, die die Indifferenz bereits bei $p = 0.99$ festlegt, besser der juristischen Forderung zu entsprechen.

1) Daß diese Zuordnung plausibel ist, findet seine Bestätigung in der Rechtspraxis selbst, in der der Grundsatz gilt: "keine Strafe ohne Verbrechen"; $y = 0$ erfüllt diese Forderung.

2) Auch diese Zuordnung ist nicht weniger plausibel, wird doch oft in der Rechtspraxis proklamiert, daß die "Nicht-Bestrafung eines Unschuldigen" das höchste juristische Gut sei.

3) Dieses Ergebnis wird nicht davon beeinflusst, ob wir x anstelle von 1 irgendeine andere Zahl zuordnen; mit $x = 0.5$ z.B. ist $z \approx 0.49$. (Die Präferenzordnung zwischen den Maßnahmen bleibt erhalten; dies läßt sich leicht zeigen, vgl. dazu auch S.274, Fn.1 dieser Arbeit).

Drückt man also den juristischen Grundsatz "mit an Sicherheit grenzender Wahrscheinlichkeit" numerisch etwa durch den Wahrscheinlichkeitswert 0.99 aus und stellt man an die Nützlichkeitsmatrix noch einige andere mit der Rechtspraxis nicht nur konsistente, sondern von ihr sogar geforderte Bedingungen, so gelangen wir in die merkwürdige Situation, daß für uns - wenn wir den Beweis mit der Wahrscheinlichkeit von 0.99 retten wollen - die Konsequenz "keine Bestrafung für die Straftat" fast ebenso wünschenswert sein muß wie die Konsequenz "Bestrafung für die Straftat". Der einzige Ausweg scheint hier die Preisgabe der Forderung nach der Beweisgrenze von 0.99 für den strafbaren Sachverhalt zu sein. Dies bedeutet allerdings nicht, daß der Sachverhalt nun mit einer geringeren Wahrscheinlichkeit zu beweisen wäre. Im Gegenteil: Der Richter muß sich nach wie vor bemühen, den Sachverhalt mit einer hohen (bzw. in seinen Augen ausreichenden) Wahrscheinlichkeit festzustellen. Ist die Wahrscheinlichkeit für den Sachverhalt z.B. wie oben 0.99, so wird man etwa mit der (plausiblen) GM

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{(die Beweisgrenze liegt hier bei der Wahrscheinlichkeit 0.5)}$$

auch M_1 (die Bestrafung des Täters im obigen Beispiel) ableiten.

Um das obige Argument weiter zu verstärken, wollen wir im folgenden verschiedene Beweisgrenzen hypothetisch annehmen und überprüfen, wie sich die Nützlichkeitsmatrix entsprechend dieser Beweisgrenzen verändert.

Seien die Tatbestände, Maßnahmen, die Konsequenzenmatrix und die Nützlichkeitsmatrix mit den Nützlichkeiten "x", "y", "z", "w" wie früher ($y = 0$, $w = 1$, x beliebig, in folgendem jedoch $= 1$).

$$p = 1$$

Die folgenden Bedingungen gelten:

$$M_1 \succ M_2 \leftrightarrow 1 \cdot x \geq 1 \cdot z + 0$$

$$M_2 \succeq M_1 \leftrightarrow 1 \cdot x \leq 1 \cdot z + 0$$

Daraus folgt: $z = 1$ und die Nützlichkeitsmatrix sieht so aus:

$$\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$$

$$p = 0.7$$

Es gelten die folgenden Bedingungen:

$$M_1 \succ M_2 \leftrightarrow 0.7 \cdot x \geq 0.7 \cdot z + 0.3$$

$$M_2 \succ M_1 \leftrightarrow 0.7 \cdot x \leq 0.7 \cdot z + 0.3$$

Für z gilt also: $z = 0.5\bar{7}$.

Die Nützlichkeitsmatrix sieht wie folgt aus:

$$\begin{bmatrix} 1 & 0 \\ 0.57 & 1 \end{bmatrix}$$

$$p = 0.3$$

Es gelten die folgenden Bedingungen:

$$M_1 \succ M_2 \leftrightarrow 0.3 \cdot x \geq 0.3 \cdot z + 0.7$$

$$M_2 \succ M_1 \leftrightarrow 0.3 \cdot x \leq 0.3 \cdot z + 0.7$$

Für z gilt: $z = -1.3\bar{3}$.

Die Nützlichkeitsmatrix sieht so aus:

$$\begin{bmatrix} 1 & 0 \\ -1.33 & 1 \end{bmatrix}$$

D.h. man muß die Konsequenz "Freispruch für den Täter" ziemlich negativ bewerten, um die Verurteilung zu erhalten. Die Bewertung der Konsequenz "Verurteilung des Nichttäters" fällt demgegenüber viel positiver aus (Wie wir sehen, müssen wir die obige Einschränkung der Nützlichkeiten auf das Intervall $[0,1]$ aufgeben. Allerdings können wir die obige Matrix wiederum in eine äquivalente positiv lineare transformieren).

Die obige Matrix sieht transformiert in eine äquivalente positive Nützlichkeitsmatrix wie folgt aus:

$$\begin{bmatrix} 1 & 0.58 \\ 0 & 1 \end{bmatrix}$$

An der Aussage dieser Matrix ändert sich gegenüber der früheren nichts, nur wird jetzt, anstatt der Konsequenz "Freispruch für den Täter" einen negativen Nutzenwert zuzuordnen, der Konsequenz "Verurteilung für den Nicht-Täter" ein positiver Nutzenwert zugeordnet (Wir müssen somit unsere ursprüngliche Annahme: $y = 0$ korrigieren).

$$p = 0.5$$

Die folgenden Bedingungen gelten:

$$M_1 \succ M_2 \leftrightarrow 0.5 \cdot x \geq 0.5 \cdot z + 0.5$$

$$M_2 \succ M_1 \leftrightarrow 0.5 \cdot x \leq 0.5 \cdot z + 0.5$$

z ist somit $= 0$.

Die Nützlichkeitsmatrix:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Das ist die uns bereits bekannte Matrix GM.

$$p = 0.1.$$

Es gelten die folgenden Bedingungen:

$$M_1 \geq M_2 \leftrightarrow 0.1 \cdot x \geq 0.1 \cdot z + 0.9$$

$$M_2 \geq M_1 \leftrightarrow 0.1 \cdot x \leq 0.1 \cdot z + 0.9$$

Für z gilt: $z = -8$.

Die Nützlichkeitsmatrix hat positiv linear transformiert folgende Gestalt:

$$\begin{bmatrix} 1 & 8/9 \\ 0 & 1 \end{bmatrix}$$

Anhand der obigen Untersuchung können wir folgendes feststellen: Hebt man die Beweisgrenze für den strafbaren Sachverhalt (d.i. der Wahrscheinlichkeitswert, der als Voraussetzung für die entscheidungstheoretische Beurteilung genommen wird) über 0.5 an, so strebt - unter einfachen plausiblen Bedingungen - die Nützlichkeitsmatrix gegen eine Auffassung, in der es offensichtlich nicht mehr auf die Bestrafung ankommt; senkt man dagegen die Beweisgrenze unter 0.5, so strebt die Nützlichkeitsmatrix gegen eine Konzeption, in der es nicht mehr darauf ankommt, daß ein Nicht-Täter nicht bestraft wird oder, was äquivalent ist, in welcher die Nicht-Bestrafung des Täters am negativsten und die Bestrafung des Täters am positivsten bewertet wird.

Liegt hingegen die Beweisgrenze bei $p = 0.5$ (mit der Annahme, daß $M_1 \geq M_2$, wenn $p \geq 0.5$ und $M_2 \geq M_1$, wenn $p < 0.5$), so erhalten wir eine Nützlichkeitsmatrix, in der sowohl die Konsequenz "Strafe für einen Täter" wie auch die Konsequenz "keine Strafe für einen Nicht-Täter" am höchsten bewertet werden und die Konsequenzen "Strafe für einen Nicht-Täter" und "keine Strafe für einen Täter" am niedrigsten, also eine Nützlichkeitsmatrix, wie wir sie der Gerechtigkeitsauffassung zugrunde legten (nach dieser Auffassung besteht das fundamentale juristische Ziel in der Wiederherstellung des Grund-

zustandes, d.h. des sozialen Gleichgewichts, vgl. S.330ff dieser Arbeit).

Legt man andererseits einer entscheidungstheoretischen Beurteilung die Gerechtigkeitsauffassung zugrunde, ohne dabei auf die Beweisgrenze zu achten, so erhält man wiederum das gleiche Ergebnis: $M_1 \geq M_2$ für $p \geq 0.5$ und $M_2 > M_1$ für $p < 0.5$. In der Verbindung mit der Gerechtigkeitsauffassung erhält somit die Beweisgrenze von 0.5 und damit das Beweiskriterium "mit überwiegender Wahrscheinlichkeit" (> 0.5) im einfachsten Fall der 2×2 -Konsequenzenmatrix eine fundamentale Bedeutung.

Außer der Matrix GM haben alle anderen abgeleiteten Nützlichkeitsmatrizen einen eher willkürlichen Charakter: Es ist kaum anzunehmen, daß jemand von vornherein auf die Idee käme, seiner Entscheidungsbildung z.B. die Nützlichkeitsmatrix

$$\begin{bmatrix} 1 & 0 \\ 0.99 & 1 \end{bmatrix}$$

zugrunde zu legen, auch wenn ihm diese auf jeden Fall den Beweis mit der sog. "an Sicherheit grenzenden Wahrscheinlichkeit" (hier: 0.99) garantiert. Geht man hingegen (ohne Einschränkung bzgl. der Beweisgrenze für den strafbaren Tatbestand) von allen möglichen strafrechtlichen Zielvorstellungen aus, so garantiert einzig und allein die Gerechtigkeitsauffassung in jedem Fall, daß diejenige Maßnahme als die beste abgeleitet wird, die dem wahrscheinlicheren Sachverhalt entspricht. Keine der anderen Zielsetzungen, von denen jede durch verschiedene Nützlichkeitsmatrizen repräsentiert werden kann, garantiert diese grundlegende Bedingung.

3.3. Das Beweiskriterium "mit kumulativ überwiegender Wahrscheinlichkeit"

Wie die obigen Ausführungen uns zeigten, erfährt das Beweiskriterium "mit überwiegender Wahrscheinlichkeit" (d.h. hier: mit der Wahrscheinlichkeit > 0.5) im Fall einer 2×2 -Konsequenzenmatrix eine formale, in Verbindung mit der Gerechtigkeitsauffassung darüberhinaus die fundamentale, Rechtfertigung.

Betrachten wir nun die eingangs auf S.342 besprochene Entscheidungssituation, in der 5 Tatbestände $T_1, \dots, T_5$ in Frage kommen. Wir nehmen folgende Wahrscheinlichkeitswerte als gegeben an:

$P(T_1) = 0.26$, $P(T_2) = 0.24$, $P(T_3) = 0.24$, $P(T_4) = 0.25$, $P(T_5) = 0.01$.

In der obigen Wahrscheinlichkeitsverteilung besitzt keiner der Tatbestände eine Wahrscheinlichkeit größer als 0.5. Würden wir die Nützlichkeitsmatrix GM ohne Einschränkung bzgl. der Beweisgrenze auf die obige Entscheidungssituation anwenden, so erhielten wir die dem Tatbestand T_1 ("Raub") entsprechende Maßnahme M_1 ("nach §250 StGB (wegen schweren Raubs)") als die beste Maßnahme (obwohl $P(T_2 \vee T_3 \vee T_4 \vee T_5) = 0.74 > P(T_1) = 0.26$ ist, d.h. die Wahrscheinlichkeit dafür, daß einer der minder schweren Tatbestände vorliegt, ist größer als die Wahrscheinlichkeit für den Tatbestand, der einzeln betrachtet, am wahrscheinlichsten ist). Mit der Annahme des Beweiskriteriums "mit überwiegender Wahrscheinlichkeit" im Sinne von " > 0.5 für den strafbaren Tatbestand" würden wir aber die dem Tatbestand T_5 ("kein Diebstahl") entsprechende Maßnahme M_5 ("Freispruch") ableiten (da keiner der strafbaren Tatbestände in diesem Sinne "überwiegt") trotz der Wahrscheinlichkeit von 0.99 dafür, daß einer der strafbaren Tatbestände $T_1, \dots, T_4$ vorliegt.

Wie wir sehen, liefert sowohl die uneingeschränkte Anwendung des Grundmodells wie auch die Forderung nach der Wahrscheinlichkeit von " > 0.5 für den strafbaren Tatbestand" in einer komplexeren Entscheidungssituation unbefriedigende Ergebnisse.

Im folgenden wird daher eine Forderung aufgestellt, wonach der strafbare Sachverhalt nicht mit "überwiegender Wahrscheinlichkeit", sondern zumindest mit "kumulativ überwiegender Wahrscheinlichkeit" festgestellt werden soll. (Was das genau heißt, wird unten erläutert.)

tert). Diese Forderung wird im Postulat JP2 festgehalten. Anschließend geben wir eine ausführliche Rechtfertigung für dieses Postulat.

Vorher sei - um Mißverständnisse zu vermeiden - noch einmal darauf hingewiesen, daß es hier nicht um die Festlegung des Wahrscheinlichkeitswertes für den Sachverhalt an sich geht, sondern um eine solche der Beweisgrenze, die für die entscheidungstheoretische Beurteilung nötig ist. Auch dann, wenn (wie bereits auf S.346 erwähnt) die Beweisgrenze im 2x2-Fall bei 0.5 liegt, so daß das Beweiskriterium "mit überwiegender Wahrscheinlichkeit" (hier im Sinn von > 0.5) gilt¹⁾, bedeutet das nicht, daß hier der Richter aufgefordert wird, mit einer geringeren Wahrscheinlichkeit, als er vielleicht erwartet, den Sachverhalt zu beurteilen. Der Richter muß nach wie vor den Sachverhalt solange prüfen, bis er der Ansicht ist, daß ihm die Wahrscheinlichkeit für diesen Sachverhalt für die nachfolgende rationale Entscheidungsbildung ausreicht (diese kann bei ihm subjektiv z.B. bei 0.99 oder aber auch unter 0.5 liegen).

Postulat JP2 besagt also, daß der strafbare Tatbestand, der der juristischen Entscheidungsbildung zugrunde gelegt wird, mit kumulativ überwiegender Wahrscheinlichkeit festgestellt werden soll; es besagt aber nicht, welchen genauen Wert diese Wahrscheinlichkeit haben sollte. Wir bezeichnen daher diese Forderung als minimale Forderung an den strafbaren Tatbestand.

Wir formulieren das Postulat JP2 in allgemeiner Form für n Tatbestände und entsprechend für r , $r \geq n$, Maßnahmen.

Zuerst einige Begriffserklärungen:

"Der Tatbestand T_i ist mit überwiegender Wahrscheinlichkeit festgestellt" bedeutet im folgenden, daß T_i wahrscheinlicher ist als jeder andere (einzelne) Tatbestand, d.h. $P(T_i) > P(T_j)$, $j = 1, \dots, n$, $i \neq j$.

"Der Tatbestand T_i ist mit kumulativ überwiegender Wahrscheinlichkeit festgestellt" bedeutet, daß die Summe der Wahrscheinlichkeiten der

1) Die beiden Forderungen "mit überwiegender Wahrscheinlichkeit" und "mit kumulativ überwiegender Wahrscheinlichkeit" stimmen für den 2x2-Fall überein. Dazu mehr auf S.355.

Tatbestände $T_1, \dots, T_i$ größer als 0.5 ist, d.h.

$$\sum_{j=1}^i P(T_j) > 0.5, i \leq n. \text{ Für } i = n \text{ ist } \sum_{j=1}^n P(T_j) = 1^{1)} \\ (\text{für alle } k > i \text{ gilt somit: } \sum_{j=1}^k P(T_j) > 0.5).$$

JP2 (Forderung nach kumulativ überwiegender Wahrscheinlichkeit)²⁾

Seien $T_1, \dots, T_n$ Tatbestände³⁾, die ihrer Schwere nach geordnet sind, d.h. T_1 sei der schwerste Tatbestand, T_n der leichteste⁴⁾. $M_1^1, \dots, M_r^n$, $r \geq n$, seien $T_1, \dots, T_n$ entsprechende Maßnahmen, so daß M_1^1 die schwerste Maßnahmen, M_r^n die leichteste Maßnahme ist. Weiter sei $P(T_1) = \max \{P(T_j)\}$, $j = 1, \dots, n$. Dann gilt:

1. Fall $P(T_1) > 0.5$, dann gilt $M_k^s \succ M_1^t \leftrightarrow E_u(M_k^s) \geq E_u(M_1^t)$ für $s, t \in \{1, \dots, n\}$.
2. Falls $P(T_1) \leq 0.5$, dann gilt: falls $\sum_{j=1}^m P(T_j) > 0.5$ und $\sum_{j=1}^{m-1} P(T_j) \leq 0.5$, $m \leq n$, dann $M_k^s \succ M_1^t \leftrightarrow E_u(M_k^s) \geq E_u(M_1^t)$ für $s, t \in \{m, \dots, n\}$.

Falls $n = 2$ und $r = n$, haben wir nur zwei Tatbestände T_1 und T_2 mit $P(T_2) = 1 - P(T_1) = P(\neg T_1)$, wobei es sich bei T_1 und T_2 um zwei strafbare Tatbestände oder um einen strafbaren Tatbestand und seine Negation handeln kann. M_1 und M_2 sind die T_1 und T_2 entsprechenden Maßnahmen (falls T_1 ein strafbarer Tatbestand ist und T_2 seine Negation, dann ist M_1 eine tatschuldangemessene Strafe und M_2 Freispruch). Nach JP2 gilt dann:

1) Eine andere Formulierung für den gleichen Zusammenhang: Die Summe der Wahrscheinlichkeiten für i Tatbestände (einschließlich T_1) ist größer als die Wahrscheinlichkeit für die restlichen $n-i$ Tatbestände, d.h. formal: $\sum_{j=1}^i P(T_j) > \sum_{j=i+1}^n P(T_j)$.

2) Im folgenden werden wir voraussetzen, daß keinem der Tatbestände die Wahrscheinlichkeit 0 zukommt. Diese Voraussetzung soll verhindern, daß eine Maßnahme gewählt wird, die einem "praktisch unmöglichen" Ereignis korrespondiert.

3) Es kann sich bei $T_1, \dots, T_n$ sowohl um "strafbare+nicht strafbare" wie auch nur um "strafbare" Tatbestände handeln.

4). Falls mehrere nicht strafbare Tatbestände $T_{n'}, \dots, T_p$ ($n' = n$) vorhanden sind (auf diese folgt Freispruch), so ist es sinnvoll, anstatt dieser den Tatbestand T_n zu betrachten, wobei gilt: $P(T_n) = \sum_{j=n'}^p P(T_j)$.

- 1) Falls $P(T_1) > P(T_2)$, dann gilt: $M_1 \succ M_2 \leftrightarrow E_u(M_1) \geq E_u(M_2)$
 und $M_2 \succ M_1 \leftrightarrow E_u(M_2) \geq E_u(M_1)$.
 2) Falls $P(T_2) \geq P(T_1)$, dann gilt $M_2 \succ M_1$ ¹⁾.

Erläuterung:

JP2 besagt, daß der strafbare Tatbestand mit kumulativ überwiegender Wahrscheinlichkeit festgestellt werden muß, um für die entscheidungstheoretische Beurteilung in Betracht gezogen zu werden.

Falls es nur zwei Tatbestände (den strafbaren Tatbestand und seine Negation) und nur zwei Maßnahmen (Strafe und Freispruch) gibt, so muß die Wahrscheinlichkeit des strafbaren Tatbestands überwiegen: erst dann kann man entscheidungstheoretisch (rational) bestimmen, welche Maßnahme zu wählen ist. Falls die Negation des strafbaren Tatbestandes mit überwiegender bzw. gleicher Wahrscheinlichkeit festgestellt ist, muß Freispruch gewählt werden. Wie wir sehen, stimmen die beiden Forderungen "mit überwiegender Wahrscheinlichkeit" und "mit kumulativ überwiegender Wahrscheinlichkeit" für nur zwei Tatbestände miteinander überein.

Falls mehrere Tatbestände und dementsprechend mehrere Maßnahmen vorhanden sind (wobei die Anzahl der Maßnahmen die der Tatbestände übersteigen kann) unterscheiden wie wiederum zwei Fälle:

1. Fall: Wenn ein Tatbestand die Summe der Wahrscheinlichkeiten aller übrigen Tatbestände übersteigt, dann wird die Entscheidungssituation auf die Weise eingeschränkt, daß alle in der graphischen Ordnung links von diesem Tatbestand stehenden Tatbestände ausgeschlossen werden; erst dann (in einer derart modifizierten Entscheidungssituation) kann man rational entscheiden, welche Maßnahme zu wählen ist. Der Entscheidungsraum wird hier also in der Weise eingeschränkt, daß man alle Tatbestände, die dem mit überwiegender Wahrscheinlichkeit nachgewiesenen Tatbestand vorangehen (d.h. die Tatbestände, die schwerer sind als dieser Tatbestand) und die ihnen

1) $P(T) = P(\neg T) \rightarrow M_2 \succ M_1$ ist übrigens ein formales Analogon für den Grundsatz "in dubio pro reo", der besagt, daß, falls der Richter unentschieden ist, ob der Tatbestand oder seine Negation gilt, er zu Gunsten des Angeklagten entscheiden soll.

entsprechenden Maßnahmen nicht mehr in Betracht zieht. Dadurch wird gewährleistet, daß zwar keine härtere Strafe (als die dem mit überwiegender Wahrscheinlichkeit nachgewiesenen Tatbestand tat-schuldangemessene Maßnahme), dagegen aber evtl. eine mildere Strafe gewählt werden kann. Wie wir sehen, stimmen auch hier (d.h. falls die Wahrscheinlichkeit für den strafbaren Tatbestand > 0.5 ist) die beiden Forderungen "mit überwiegender Wahrscheinlichkeit" und "mit kumulativ überwiegender Wahrscheinlichkeit" überein.

2.Fall: Wenn ein strafbarer Tatbestand zwar überwiegt, jedoch nicht wahrscheinlicher ist als die Summe der Wahrscheinlichkeiten aller übrigen Tatbestände, dann wird die Entscheidungssituation in der Weise eingeschränkt, daß alle in der graphischen Ordnung links stehenden Tatbestände, deren kumulative Wahrscheinlichkeit 0.5 nicht überschreitet, ausgeschlossen werden. D.h. der Tatbestand T_m , der in der eingeschränkten Entscheidungssituation in der graphischen Reihenfolge jetzt am meistens links steht, ist derjenige, bei dem die Summe der Wahrscheinlichkeiten der Tatbestände T_1 bis T_m , $m \leq n$, zum ersten Mal die Wahrscheinlichkeit 0.5 überschreitet. Das verhindert die Ableitung der Entscheidung entsprechend einem Tatbestand (auch wenn dieser der wahrscheinlichste ist), wenn die Wahrscheinlichkeit für das Vorliegen dieses oder eines der schwereren Tatbestände (also die kumulative Wahrscheinlichkeit für diesen Tatbestand) ≤ 0.5 ist.

JP2 kann vereinfacht werden, indem man die Unterscheidung zwischen den Fällen 1 und 2 aufgibt, und stattdessen fordert:

Falls $\sum_{j=1}^m P(T_j) > 0.5$ und $\sum_{j=1}^{m-1} P(T_j) \leq 0.5$, $m \leq n$, dann gilt:

$$M_k^s \rightarrow M_1^t \leftrightarrow E_u(M_k^s) \geq E_u(M_1^t) \text{ für } s, t \in \{m, \dots, n\}.$$

Mit dieser Forderung werden die beiden Fälle 1 und 2 aus JP2 erfüllt. Die Formulierung von JP2 wird teilweise aus Gründen der Übersicht, teilweise aus denen der Didaktik gewählt, um den Unterschied zwischen der Forderung nach "überwiegender Wahrscheinlichkeit (im Sinne von > 0.5)" und "kumulativ überwiegender Wahrscheinlichkeit" zu veranschaulichen. Im ersten Fall fällt, wie wir bereits gesehen haben, die Forderung nach kumulativ überwiegender Wahrscheinlichkeit mit der nach überwiegender Wahrscheinlichkeit zusammen. Erst im zweiten Fall ist die Unterscheidung zwischen den beiden Beweiskriterien wichtig.

Zur Veranschaulichung von JP2:

(die jeweiligen maximalen Wahrscheinlichkeiten sind unterstrichen; mit vertikalen Strichen ist die neue (eingeschränkte) Tatbestandsmenge gekennzeichnet, d.h. alle links von dem so gekennzeichneten Tatbestand stehenden Tatbestände werden außer Acht gelassen).

I. strafbare + nicht strafbare Tatbestände

	Mord	Totschlag	minder schw. Fall des Totschlags	Notwehr
1.	<u>0.6</u>	0.2	0.1	0.1
	0.2	<u>0.6</u>	0.1	0.1
	0.1	0.1	<u>0.6</u>	0.2
	0.1	0.2	0.1	<u>0.6</u>
2.	<u>0.4</u>	<u>0.2</u>	0.1	0.3
	0.1	<u>0.4</u>	<u>0.2</u>	0.3
	0.1	<u>0.5</u>	0.1	0.3
	0.1	0.1	<u>0.4</u>	<u>0.4</u>
	0.3	0.2	<u>0.1</u>	<u>0.4</u>
	<u>0.25</u>	<u>0.25</u>	<u>0.25</u>	<u>0.25</u>
	0.1	0.1	0.3	<u>0.5</u>

II strafbare Tatbestände

	Mord	Totschlag	minder schw. Fall des Totschlags
1.	<u>0.6</u>	0.1	0.3
	0.2	<u>0.6</u>	0.2
	0.3	0.1	<u>0.6</u>
2.	<u>0.5</u>	<u>0.1</u>	0.4
	<u>0.4</u>	<u>0.2</u>	<u>0.4</u>
	0.1	0.4	<u>0.5</u>
	<u>0.3</u>	<u>0.3</u>	<u>0.3</u>
	0.1	<u>0.5</u>	0.4

Beispiel:

Angenommen, vier Tatbestände: "Mord" (T_1), "Totschlag" (T_2), "fahrlässige Tötung" (T_3), "Notwehr" (T_4) und fünf Maßnahmen: "lebenslange Freiheitsstrafe" (M_1^1), "Freiheitsstrafe nicht unter 5 Jahren" (M_2^2), "Freiheitsstrafe bis zu 5 Jahren" (M_3^3), "Geldstrafe" (M_4^4) und "Freispruch" (M_5^5)¹⁾ werden in Betracht gezogen. Falls $P(T_1) = P(T_2) = P(T_3) = P(T_4) = 0.25$, dann wird nach JP2 die Tatbestandsmenge auf $\{T_3, T_4\}$ und die Maßnahmenmenge auf $\{M_3^3, M_4^4, M_5^5\}$ eingeschränkt.

Sei etwa $P(T_1) = 0.4$, $P(T_2) = 0.1$, $P(T_3) = 0.3$ und $P(T_4) = 0.2$. Es gilt $P(T_1) > P(T_2)$, $P(T_1) > P(T_3)$, $P(T_1) > P(T_4)$, $P(T_1) < P(T_2) + P(T_3) + P(T_4)$, $P(T_1 \vee T_2) = 0.5$ und $P(T_1 \vee T_2 \vee T_3) = 0.8$. Also wird nach JP2 wiederum die Tatbestandsmenge auf $\{T_3, T_4\}$ und die Maßnahmenmenge auf $\{M_3^3, M_4^4, M_5^5\}$ eingeschränkt.

Einige kritische Anmerkungen:

1. Seien drei Tatbestände "Mord" (T_1), "Totschlag" (T_2) und "minder schwerer Fall des Totschlags" (T_3) gegeben und jeweils mit einer Wahrscheinlichkeit von $1/3$ nachgewiesen. Da die kumulative Wahrscheinlichkeit bei "Mord" $1/3$ und bei "Totschlag" $2/3$ beträgt, müssen wir nach JP2 die Tatbestandsmenge auf $\{T_2, T_3\}$ einschränken. Man kann sich fragen, wieso die Tatbestandsmenge nicht auf $\{T_1, T_3\}$ oder aber auf $\{T_1, T_2\}$ eingeschränkt wird, denn auch hier gilt, daß die Wahrscheinlichkeit für das Vorliegen eines der Tatbestände aus diesen Mengen $2/3$ beträgt.

Vergleicht man allerdings die drei Tatbestandsmengen $\{T_1, T_2\}$, $\{T_2, T_3\}$ und $\{T_1, T_3\}$ genauer miteinander, so zeigt sich, daß die Menge $\{T_2, T_3\}$ die leichtesten Tatbestände enthält. Nach JP2 gilt also: Haben mehrere Teilmengen der Tatbestandsmenge eine Wahrscheinlichkeit von ≥ 0.5 , so wird aus diesen diejenige (kleinste) gewählt, die die leichtesten Tatbestände enthält (d.h. diejenige, die in der graphischen Ordnung am meisten rechts steht)²⁾.

1) Der obere Index gibt den Tatbestand an, dem die Maßnahme entspricht.

2) Inhaltlich bedeutet dies das gleiche wie die (in der Formulierung von JP2 gebrauchte) Aussage, daß alle Tatbestände links in der graphischen Ordnung, deren kumulative Wahrscheinlichkeit ≤ 0.5 ist, aus der Tatbestandsmenge ausgeschlossen werden.

2. Nach JP2 werden zwar einige Tatbestände ausgeschlossen, die eingeschränkte Tatbestandsmenge kann aber trotzdem Tatbestände enthalten, die einzeln betrachtet weniger wahrscheinlich sind als die ausgeschlossenen. Man kann sich fragen, ob das gerecht ist. Betrachten wir etwa die Tatbestände "Mord" (T_1), "Totschlag" (T_2), "Notwehr" (T_3) mit den jeweiligen Wahrscheinlichkeiten 0.5, 0.4, 0.1. Nach JP2 wird "Mord" ausgeschlossen, obwohl $P(T_1) > P(T_2)$ ist. Schauen wir uns aber die Wahrscheinlichkeit für das Vorliegen von wenigstens einem der strafbaren Tatbestände an. Diese ist $P(T_1 \vee T_2) = 0.9$, d.h. die Wahrscheinlichkeit, daß einer der strafbaren Tatbestände vorliegt (oder die kumulative Wahrscheinlichkeit für T_2) ist 0.9. In dieser Situation, wo zwar die Wahrscheinlichkeit für das Vorliegen eines der strafbaren Tatbestände > 0.5 ist, keiner der strafbaren Tatbestände jedoch über 0.5 liegt, ist es aus humanitären Gründen auf jeden Fall gerechtfertigt, nach dem minder schweren Tatbestand vorzugehen (auch wenn - wie im obigen Beispiel - $P(T_1) > P(T_2)$ ist).

Wir wollen die Forderung nach "kumulativ überwiegender Wahrscheinlichkeit" noch in der qualitativen Sprechweise formulieren.

JP2* Der Tatbestand muß zumindestens mit kumulativ überwiegender Wahrscheinlichkeit festgestellt werden, d.h. genauer:

$T_1, \dots, T_n$ seien der Schwere nach geordnete Tatbestände, d.h. T_1 sei der schwerste Tatbestand, T_n der leichteste; $M_1^1, \dots, M_r^n$, $r \geq n$, seien die $T_1, \dots, T_n$ entsprechenden Maßnahmen, so daß M_1^1 die schwerste Maßnahme und M_r^n die leichteste Maßnahme ist. Weiter sei T_i der Tatbestand, der am wahrscheinlichsten ist, $i = 1, \dots, n$. Dann gilt:

1. Falls man T_i für wahrscheinlicher hält als alle anderen Tatbestände zusammengenommen, dann gilt $M_K^S \geq M_i^T$ oder $M_i^T \geq M_K^S$ für $s, t \in \{1, \dots, n\}$.
2. Falls man T_i für gleich oder weniger wahrscheinlich hält als alle anderen Tatbestände zusammengenommen, dann suche man aus der Tatbestandsmenge diejenige (kleinste) Teilmenge aus, für die erstens gilt: die Wahrscheinlichkeit für das Vorliegen eines dieser Tatbestände ist nicht kleiner als die Wahrscheinlichkeit für das Vorliegen eines der restlichen Tatbestände und die zweitens die leichtesten Tatbestände enthält, d.h. die, die in der graphischen Ordnung am meisten rechts stehen. Anschließend wähle man aus der diesen Tatbeständen entsprechenden (eingeschränkten) Maßnahmenmenge eine Maßnahme.

Erläuterung:

JP2* besagt, daß nur dann eine Maßnahme entsprechend einem Tatbestand oder eine solche, die einem leichteren als diesem Tatbestand korrespondiert, gewählt werden darf, wenn man alle schwereren Tatbestände einschließlich des betrachteten Tatbestands zusammengenommen für wahrscheinlicher hält als die leichteren Tatbestände zusammengenommen (ohne den betreffenden Tatbestand).

Rechtfertigung von JP2

Bei der Rechtfertigung von JP2 werden u.a. einige Argumente, die in der einen oder anderen Form in dieser Arbeit bereits vorgetragen wurden, noch einmal wiedergegeben.

1. An mehreren Stellen dieser Arbeit sprachen wir davon, daß statt Objektivität (im Sinne der vollständigen Übereinstimmung mit der Wirklichkeit) Rationalität die oberste Maxime bei dem juristischen Vorgehen unter Unsicherheit im weiteren Sinn (Risiko) sein muß. Dementsprechend ersetzten wir den konventionellen juristischen Überzeugungsmaßstab der "vollen Gewißheit" durch denjenigen der "Glaubwürdigkeit", d.h. der subjektiven Wahrscheinlichkeit.

2. Geht man bei der juristischen Entscheidungsbildung von einer Wahrscheinlichkeit für den Sachverhalt aus, so ist die ganze Situation entscheidungstheoretisch rekonstruierbar: Man benötigt nur noch eine Nützlichkeitsmatrix, um eine Entscheidung herzuleiten. Ein Problem taucht hier jedoch auf: Die unbeschränkte Anwendung einer entscheidungstheoretischen Konzeption auf die Sachverhalte (wie das im Savage-Modell der Fall ist) kann dazu führen, daß die dem weniger wahrscheinlichen strafbaren Sachverhalt entsprechende Entscheidung als die beste hergeleitet wird (etwa dann, wenn der Angeklagte verurteilt wird, obwohl die Wahrscheinlichkeit dafür, daß er die Tat begangen hat, kleiner ist als diejenige für das Gegenteil). Um derartige Ergebnisse zu vermeiden, muß man offenbar eine entscheidungstheoretische Konzeption in einer dafür geeigneten Weise einschränken.

3. Hier entsteht gleich die Frage: In welcher Form soll man ein Postulat formulieren, das die oben erwähnte Schwierigkeit vermeidet?

Im Falle von nur zwei Sachverhalten (dem Tatbestand und seiner Negation) ist das Problem (offensichtlich) leicht lösbar: der strafbare Tatbestand muß auf jeden Fall wahrscheinlicher sein als seine Negation (das gleiche gilt bei zwei Sachverhalten, von denen keiner die Negation des anderen ist).

Wenn mehr als zwei Tatbestände vorliegen, so kann man es auch hier mit der Forderung versuchen, daß der schwerste in Frage kommende Tatbestand auf jeden Fall wahrscheinlicher als 0.5 sein soll, um für die entscheidungstheoretische Beurteilung in Betracht gezogen zu werden. Wir haben allerdings gesehen, daß eine solche Forderung zu unplausiblen und unbefriedigenden Ergebnissen führt, etwa in der Situation, wo die Wahrscheinlichkeit für das Vorliegen eines der

strafbaren Tatbestände 0.99 beträgt und die Wahrscheinlichkeit für den nicht strafbaren Tatbestand nur 0.01 ist, wir aber trotzdem "Freispruch" wählen müssen (vgl. S.350). Auch die Forderung, daß der schwerste in Frage kommende Tatbestand lediglich wahrscheinlicher sein soll als jeder einzelne andere Tatbestand, führte zu unbefriedigenden Ergebnissen (vgl. S.350 dieser Arbeit). Um einerseits derartige unplausible und unbefriedigende Ergebnisse auszuschließen und andererseits zu vermeiden, daß eine Entscheidung entsprechend einem Tatbestand abgeleitet wird, der zusammen mit den schwereren Tatbeständen die leichteren nicht überwiegt, haben wir den Begriff der kumulativen Wahrscheinlichkeit (sowie den der graphischen Ordnung) eingeführt und verlangt: Der schwerste der in Frage kommenden Tatbestände muß derjenige sein, bei dem zum ersten Mal die kumulative Wahrscheinlichkeit größer als 0.5 ist, wobei alle links von ihm in der graphischen Ordnung stehenden (schwereren) Tatbestände außer Acht gelassen werden. Hier wird also nach der Maxime gehandelt: Wenn die Wahrscheinlichkeit dafür, daß z.B. "Mord" oder "Totschlag" vorliegt, größer als 0.5 ist, allerdings die Wahrscheinlichkeit keines dieser beiden Tatbestände 0.5 übersteigt, dann ist es besser, eine Entscheidung entsprechend "Totschlag" abzuleiten als eine entsprechend "Mord". (Beachte, daß auch hier eine Entscheidung gemäß einem weniger wahrscheinlichen Tatbestand möglich ist, nur wiegt dieser Tatbestand entweder leichter als derjenige, der mit überwiegender Wahrscheinlichkeit festgestellt ist, oder er ist am leichtesten in der Teilmenge der Tatbestandsmenge, die ihn und alle links von ihm stehenden (also schwereren) Tatbestände enthält und deren Wahrscheinlichkeit (d.h. die der Teilmenge) größer als 0.5 ist.

4. Den obigen Erwägungen haben wir in Form des Postulats JP2 Ausdruck verliehen.

Im Postulat JP2 haben wir gefordert, daß für die entscheidungstheoretische Beurteilung unabhängig von der Anzahl der Tatbestände ein strafbarer Tatbestand entweder mit einer Wahrscheinlichkeit größer der Summe der Wahrscheinlichkeiten aller anderen Tatbestände festgestellt werden muß, oder - wenn das nicht der Fall ist - die Summe der Wahrscheinlichkeiten aller schwereren Tatbestände einschließlich des betreffenden (d.h. die kumulative Wahrscheinlichkeit für diesen Tatbestand) größer sein muß als die Summe der Wahrscheinlichkeiten aller leichteren Tatbestände.

Diese Forderung an den strafbaren Tatbestand gibt somit eine notwendige Bedingung an - die wir als minimale Bedingung bezeichneten - und die immer bei der Entscheidungsbildung erfüllt sein muß.

5. Wir haben vorher JP2 sozusagen von 'unten' begründet (als minimale Forderung an den strafbaren Tatbestand). Der strafbare Tatbestand muß lediglich wahrscheinlicher sein als seine Negation bzw. kumulativ 0.5 überwiegen, um für die entscheidungstheoretische Beurteilung relevant zu sein. Diese Forderung wird somit auch dann erfüllt, wenn die kumulative Wahrscheinlichkeit für den strafbaren Sachverhalt den minimalsten Wert im Intervall $(0.5, 1]$ annimmt¹⁾. Angesichts dessen, daß auch Wahrscheinlichkeitswerte wie 0.6, 0.7 etc. diese Forderung erfüllen, entsteht das Problem, JP2 auch von 'oben', d.h. als die einzig in Frage kommende Forderung, zu rechtfertigen²⁾. Folgende Rechtfertigungsgründe können hier angeführt werden³⁾:

- Jede weitere Einschränkung des Intervalls $(0.5, 1]$ ist willkürlich, d.h. eine Wahl von 0.6, 0.7 etc. als die unterste Grenze für den Sachverhaltsbeweis kann nicht rational gerechtfertigt werden. Unter einem Wahrscheinlichkeitswert von 0.6, 0.7 etc. kann man sich außerdem wenig vorstellen (im Unterschied zur Ausdrucksform "... ist wahrscheinlicher als ...", die intuitiv klar ist und deren Gebrauch keine Schwierigkeiten bereitet).

Nachdem es keine rationale Rechtfertigung für die Wahl eines konkreten Wahrscheinlichkeitswertes als unterer Grenze für den Beweis des Tatbestandes gibt, ist es mehr als plausibel, den Grundsatz nach "kumulativ überwiegender Wahrscheinlichkeit" in Form von JP2 als die einzig sinnvolle Forderung an den strafbaren Tatbestand zu akzeptieren.

6. Was heißt aber, daß die kumulative Wahrscheinlichkeit für den strafbaren Tatbestand >0.5 sein muß? - Lediglich, daß der Entscheidende den strafbaren Tatbestand nur dann für die entscheidungstheoretische Beurteilung in Betracht zieht, wenn er mit einer größeren kumulativen

1) Ein solcher Wert kann übrigens nie praktisch angenommen werden.
2) Man könnte z.B. verlangen, daß die Summe der Wahrscheinlichkeiten aller schwereren Tatbestände + des in Frage kommenden Sachverhalts etwa über 0.9 liegt, im Beispiel auf 0.356 $P(T_1) + P(T_2) + P(T_3) < 0.9$ und $P(T_1) + P(T_2) + P(T_3) + P(T_4) > 0.9$. Nach diesem Kriterium wird die Tatbestandsmenge auf $\{T_4\}$ eingeschränkt.

3) Diese Rechtfertigungsgründe gelten nicht nur für die Forderung nach "kumulativ überwiegender Wahrscheinlichkeit", sondern auch für diejenige nach "überwiegender Wahrscheinlichkeit im Sinne von >0.5 ".

Wahrscheinlichkeit als 0.5 nachgewiesen ist. Das zieht nicht notwendigerweise die Verurteilung des Angeklagten entsprechend diesem Sachverhalt nach sich. Aber auch mit diesem Wissen, daß z.B. eine kumulative Wahrscheinlichkeit von 0.51 ihm ausreichen würde, um (z.B. nach der GM) den Angeklagten zu verurteilen, wird ein gewissenhafter Richter bestrebt sein, den Sachverhalt mit möglichst hoher Wahrscheinlichkeit zu beweisen, vergleichbar etwa der folgenden Situation: Einem Richter möge die Aussage eines Zeugen, falls er diesem Zeugen mehr glaubt als einem anderen, der das Gegenteil behauptet, genügen, den Angeklagten zu verurteilen, auch wenn der zweite Zeuge bei weitem noch nicht völlig an Glaubwürdigkeit eingebüßt hat. Ein gewissenhafter Richter dagegen wird sich bemühen, durch neue, zusätzliche Informationen die Glaubwürdigkeit des ersten Zeugen noch weiter zu bekräftigen und die des anderen noch stärker zu verringern, auch wenn er ohne solche Informationen die Verurteilung aussprechen könnte.

Wir müssen somit klar sehen, was das Postulat JP2 leistet und was nicht:

Durch JP2 wird eine Entscheidung entsprechend dem kumulativ weniger wahrscheinlichen Tatbestand vermieden und eine entscheidungstheoretische Anwendung auf die juristischen Sachverhalte in sinnvoller (rationaler) Weise garantiert. Daher ist ein solches Postulat notwendig.

Es liefert bzw. es verlangt jedoch nicht explizit einen genügend hohen Wahrscheinlichkeitswert für den strafbaren Tatbestand, der in seiner Ausdruckskraft dem Begriff von der "völligen Überzeugung" bzw. der "an Sicherheit grenzenden Wahrscheinlichkeit" annähernd entsprechen würde. Dies ist aber auch nicht nötig: Erstens variiert das, was man subjektiv unter dem Beweiskriterium "mit an Sicherheit grenzender Wahrscheinlichkeit" versteht, von Fall zu Fall und von Individuum zu Individuum¹⁾. Es gibt also keinen Grund, einen dieser

1) Wie mehrere empirische Untersuchungen zeigen (vgl. I. Teil, 2. Kap.) kann jeder beliebige Wahrscheinlichkeitswert aus dem Intervall zwischen 0.5 und 1 für ein Individuum diese "an Sicherheit grenzende Wahrscheinlichkeit" repräsentieren. Allerdings zog man in allen diesen Untersuchungen nur zwei Umstände: "schuldig"/"unschuldig" in Betracht. Das Postulat JP2 trägt also für den einfachsten Fall von nur zwei Tatbeständen diesen empirischen Untersuchungen Rechnung. Für mehr als zwei Tatbestände existieren leider keine vergleichbaren empirischen Untersuchungen. Aus theoretischer Sicht haben wir oben Bedenken dagegen vorgetragen, die Forderung nach "überwiegender Wahrscheinlichkeit im Sinne von >0.5 " auch auf solche Fälle zu übertragen, wo mehr als zwei strafbare Tatbestände vorliegen.

Werte auszuzeichnen. Dies beachtend habe ich auf S.169 vorgeschlagen, anstatt von der "an Sicherheit grenzenden Wahrscheinlichkeit" von der Wahrscheinlichkeit zu sprechen, die einem Individuum in einem konkreten Fall für die Verurteilung, hier: für die entscheidungstheoretische Beurteilung, ausreicht.

Zweitens hat der Versuch, den Grundsatz "mit an Sicherheit grenzender Wahrscheinlichkeit", repräsentiert durch den geläufigen Ausdruck 0.99, entscheidungstheoretisch zu verankern, gezeigt, daß dies zu einer sehr unplausiblen Einschränkung bzgl. der Zuordnung von Nützlichkeiten zu den Konsequenzen führt.

Schließlich soll hier noch einmal gesagt werden, daß die Forderung an den strafbaren Tatbestand für dessen entscheidungstheoretische Beurteilung keineswegs eine Verhaltensmaxime an den Richter bei seiner wahrscheinlichkeitstheoretischen Beurteilung des Tatbestandes darstellt, d.h. es wird vom Richter nicht verlangt, daß er den Tatbestand in jedem Fall oder nur mit einer kumulativ überwiegenden Wahrscheinlichkeit feststellen soll. Mit welcher Wahrscheinlichkeit der Richter den Sachverhalt feststellen will, darüber entscheidet der Richter selbst; JP2 zeigt ihm lediglich, ob diese Wahrscheinlichkeit dann auch tatsächlich für die entscheidungstheoretische Beurteilung und darüber hinaus für die Verurteilung genügt oder nicht.

Zusammenfassend läßt sich somit folgendes sagen:

Die Einschränkung einer entscheidungstheoretischen Konzeption ist notwendig. Das leistet das Postulat JP2. Eine weitere Einschränkung ist nicht möglich, weil sie willkürlich wäre. Alle anderen (im Kontext untersuchten) alternativen Formulierungen eines entsprechenden Postulates scheiden aus, weil sie zu unbefriedigenden Ergebnissen führen würden. Der Richter muß sich somit - mit Wissen, daß ihm auch eine kumulative Wahrscheinlichkeit von nur 0.51 für den strafbaren Sachverhalt genügen würde, um diesen entscheidungstheoretisch zu beurteilen und evtl. eine Entscheidung entsprechend diesem Sachverhalt zu fällen - selbst bemühen, den Sachverhalt mit einem möglichst hohen Wahrscheinlichkeitswert zu beweisen.

3.4 Reduktion der entscheidungstheoretischen Beurteilung auf die wahrscheinlichkeitstheoretische aufgrund der Gerechtigkeitsauffassung

Kehren wir zu der auf S.330ff dieser Arbeit entwickelten Gerechtigkeitsauffassung zurück. Wir legten dieser Auffassung die Vorstellung zugrunde, daß eine Straftat eine (negative) Veränderung des sozialen Gleichgewichts oder des juristischen Status Quo darstellt und daß die (gerechte, d.h. tatschuldangemessene) Strafe diesen Status Quo formal wiederherstellt.

Die formale Wiederherstellung des juristischen Status Quo bezeichnen wir als das fundamentale juristische Ziel. Die Gerechtigkeitsauffassung, innerhalb deren dieses Ziel verfolgt wird, ist dementsprechend die fundamentale Theorie über Sinn und Zweck der Strafe. Formal gesehen wird die Gerechtigkeitsauffassung durch eine sehr einfache Nützlichkeitsmatrix repräsentiert, in der jeder angemessenen Behandlung für einen entsprechenden Tatbestand der höchste Nutzenwert 1 und allen anderen nicht angemessenen Behandlungen der niedrigste Nutzenwert 0 zugewiesen wird. Wir haben außerdem festgestellt, daß einzig und allein die Gerechtigkeit als Zielsetzung den in der Rechtspraxis geforderten Grundsatz der Gleichbehandlung vor dem Gesetz formal (wegen der Eindeutigkeit der Nutzenzuordnung) garantiert, da auf individuelle Unterschiede in keiner Weise Rücksicht genommen wird.

Angesichts der Vielfalt an strafrechtlichen Auffassungen bzgl. Ziel und Zweck der Strafe erscheint eine von vornherein getroffene Festlegung auf eine bestimmte Nützlichkeitsmatrix (in diesem Fall die GM, die der Gerechtigkeitsauffassung zugrunde liegt) allerdings wenig wünschenswert, denn es werden danach keine anderen für die Rechtspraxis wichtigen Ziele wie Resozialisation, Prävention usw. und deren mögliche Zwischenformen mehr berücksichtigt.

Außerdem sollte auch der Tatsache Rechnung getragen werden, daß darüber, welcher strafrechtliche Aspekt im Vordergrund steht, die jeweilige Rechtsprechung (konkret: der beurteilende Richter) entscheidet.

Hier wollen wir jedoch den Gerechtigkeitsgedanken weiter verfolgen und die Schlußfolgerungen betrachten, die die entscheidungstheoretische Beurteilung aufgrund der Gerechtigkeitsauffassung für die wahrscheinlichkeitstheoretische Beurteilung des Sachverhalts impliziert.

Legt man im einfachsten (2×2 -) Fall der entscheidungstheoretischen Beurteilung die Gerechtigkeitsmatrix GM zugrunde, so reduziert sich, wie wir auf S.348f gesehen haben, die erstere auf die wahr-scheinlichkeitstheoretische Beurteilung des Sachverhalts: Für $p > 0.5$ wird "Strafe" und für $p < 0.5$ "keine Strafe" abgeleitet. Nach JP1 wird auch für $p = 0.5$ "keine Strafe" abgeleitet. D.h. aus allen strafrechtlichen Auffassungen bzgl. des Sinns und Zwecks der Strafe garantiert nur die Gerechtigkeitsauffassung die Ableitung der besten Maßnahme entsprechend dem wahrscheinlicheren Tatbestand.

Stehen allerdings mehr als zwei Tatbestände zur Auswahl, so ist zwar auch hier die Ableitung nach dem wahrscheinlicheren Tatbestand gewährleistet¹⁾, doch geraten wir hier in die (bereits kritisierte) unbefriedigende Lage, eine Maßnahme als die beste entsprechend dem Tatbestand abzuleiten, der zwar wahrscheinlicher ist als jeder andere Tatbestand, allgemein gesehen jedoch ziemlich unwahrscheinlich ist (etwa bei der folgenden Wahrscheinlichkeitsverteilung: 0.3, 0.1, 0.1, 0.25, 0.25; nach der GM wird die Maßnahme entsprechend dem ersten, der am schwersten wiegt, abgeleitet). Wir müssen somit das Postulat JP2 auch dann anwenden, wenn wir von der Gerechtigkeitsauffassung ausgehen, um derartige unbefriedigende Ergebnisse zu verhindern. Im obigen Beispiel reduziert sich nach JP2 die Tatbestandsmenge auf die beiden letzten Tatbestände T_4 und T_5 . Es gilt somit $E_u(M_4) = p_4 = 0.25$, $E_u(M_5) = p_5 = 0.25$. Nach JP1 ist also M_5 die beste Maßnahme.

Das Ergebnis, daß die entscheidungstheoretische Beurteilung auf die wahr-scheinlichkeitstheoretische reduzierbar ist, wenn man die fundamentale Zielsetzung verfolgt, wollen wir in Form eines Postulats festhalten, das die Gestalt der Nützlichkeitsmatrix festlegt, die der Gerechtigkeitsauffassung zugrunde liegt. Somit wird nicht nur - wie es bislang der Fall war - plausibel begründet, warum die Gerechtigkeitsmatrix eine bestimmte Gestalt hat, sondern auch normativ (explizit) gefordert, daß die Gerechtigkeitsmatrix diese Gestalt haben muß.

1) Die entscheidungstheoretische Beurteilung reduziert sich, falls man ihr die Gerechtigkeit als Zielsetzung bzw. die Matrix GM zugrunde legt, auf die Wahrscheinlichkeitsfeststellung des Tatbestands. Denn es gilt mit der GM: $E_u(M_1) = p_1$, $E_u(M_2) = p_2, \dots, E_u(M_n) = p_n$.

JP3 (Gerechtigkeitspostulat). Seien $T_1, \dots, T_n$ Tatbestände; $M_1, \dots, M_n$ die ihnen entsprechenden Maßnahmen und R_{ij} die Konsequenzen (M_i, T_j) , $i, j = 1, \dots, n$. Dann gilt:

Falls die Gerechtigkeit als Zielsetzung der entscheidungstheoretischen Beurteilung eines juristischen Falls zugrunde liegt, dann ist die Nützlichkeitsmatrix wie folgt bestimmt:

$$u(R_{ij}) = \begin{cases} 1, & \text{für } i = j \\ 0, & \text{für } i \neq j \end{cases}$$

JP3 besagt, daß die Nützlichkeitsmatrix von einer bestimmten Form sein muß, wenn man die Annahme macht, daß man als Zielsetzung Gerechtigkeit verfolgt. Diese Annahme wollen wir im folgenden Gerechtigkeitsannahme nennen.

In JP3 entsprach die Anzahl der Maßnahmen der der Tatbestände; die Nützlichkeitsmatrix war daher eine $n \times n$ -Matrix. Von früher her wissen wir aber, daß die Anzahl der Maßnahmen größer sein kann als die der Tatbestände (wenn für ein und denselben Tatbestand mehrere Maßnahmen, etwa Gefängnisstrafe, Geldstrafe, Einweisung in eine Anstalt etc. in Frage kommen). Die Gerechtigkeitskonzeption verlangt in diesem Fall, daß nur eine davon als angemessen bzw. als tatschuldangemessen gilt und als solche den Nützlichkeitswert 1 erhält, allen anderen dagegen als nicht tatschuldangemessenen Maßnahmen der Nützlichkeitswert 0 zugewiesen wird. Eine $r \times n$ -Matrix, die $r-n$ Nullzeilen enthält, kann bekanntlich auf eine $n \times n$ -Matrix reduziert werden. Dies wollen wir in einer Erweiterung von JP3 festhalten.

JP3* (Erweiterung von JP3). Seien $T_1, \dots, T_n$ Tatbestände; $M_1^1, \dots, M_r^n$, $r \geq n$, die darauf folgenden Maßnahmen und R_{ij} die Konsequenzen (M_i^j, T_j) , $i = 1, \dots, r$, $j = 1, \dots, n$. Dann gilt:

Mit der Gerechtigkeitsannahme erhalten nur die tatschuldangemessenen Maßnahmen den Nützlichkeitswert 1, alle anderen hingegen den Nützlichkeitswert 0.

Außerdem gilt: Man kann $r-n$ Nullzeilen aus der Nützlichkeitsmatrix entfernen. Die so erhaltene Matrix hat die Gestalt:

$$u(R_{ij}) = \begin{cases} 1, & \text{für } i = j \\ 0, & \text{für } i \neq j. \end{cases}$$

Satz: JP2 und JP3 seien erfüllt. Außerdem gelte die Gerechtigkeitsannahme. ($T_1, \dots, T_n$ seien disjunkte und erschöpfende Tatbestände; $M_1, \dots, M_n$ die ihnen entsprechenden Maßnahmen; T_i sei der Tatbestand, der mit überwiegender Wahrscheinlichkeit festgestellt ist). Dann gilt:

$$(1) P(T_i) > 0.5 \leftrightarrow M_i > M_j \text{ für } j = i+1, \dots, n;$$

$$(2) P(T_i) < 0.5 \text{ und } \sum_{j=1}^m P(T_j) > 0.5 \text{ und } \sum_{j=1}^{m-1} P(T_j) < 0.5 \\ \max\{P(T_k)\} \leftrightarrow M_k \succcurlyeq M_j \text{ für } k \geq m, j = m, \dots, n.$$

(Beweis trivial).

Dieser Satz besagt, daß unter der Gerechtigkeitsannahme die entscheidungstheoretische Beurteilung des Falls (d.h. die Wahl der Maßnahme) einzig aus der wahrscheinlichkeitstheoretischen Feststellung des Tatbestandes folgt. In diesem Fall braucht man nur die Wahrscheinlichkeitswerte für die Tatbestände zu kennen, um gleich die richtige Maßnahme zu erhalten.

Der obige Satz kann vereinfacht werden, wenn die Unterscheidung in die Fälle (1) und (2) aufgegeben wird:

Satz*: JP2 und JP3 seien erfüllt. Außerdem gelte die Gerechtigkeitsannahme. $T_1, \dots$ seien disjunkte und erschöpfende Tatbestände; $M_1, \dots, M_n$ die ihnen entsprechenden Maßnahmen. Dann gilt:

$$\sum_{j=1}^m P(T_j) > 0.5 \text{ und } \sum_{j=1}^{m-1} P(T_j) < 0.5 \rightarrow \max\{P(T_k)\} \leftrightarrow M_k \succcurlyeq M_j \text{ für } k \geq m, j = m, \dots, n.$$

Beispiele:

Seien drei Tatbestände T_1 , T_2 und T_3 mit den Wahrscheinlichkeiten 0.01, 0.98 und 0.01 gegeben. Ferner sei die Annahme gemacht, daß von der Gerechtigkeitsauffassung ausgegangen wird. Nach JP2 reduziert sich die Tatbestandsmenge auf $\{T_2, T_3\}$ und nach JP3 erhalten wir die T_2 entsprechende Maßnahme als die beste.

Sei nun die folgende Wahrscheinlichkeitsverteilung $P(T_1) = 0.5$, $P(T_2) = 0.2$, $P(T_3) = 0.3$ gegeben. Mit der Gerechtigkeitsannahme erhalten wir nach JP2 und JP3 die T_3 entsprechende Maßnahme als die beste.

Wie bereits erwähnt, haben wir angesichts der Vielfalt an Auffassungen bzgl. Ziel und Zweck der Strafe die rationale juristische Entscheidungstheorie möglichst allgemein konzipiert d.h. wir wollten uns

nicht von vornherein auf eine bestimmte, in unserem Fall die Gerechtigkeitsauffassung, beschränken. Aus diesem Grund formulierten wir auch das Postulat JP3 lediglich in der Konditionalform. So folgt zwar aus JP1 - JP3 mit der Gerechtigkeitsannahme die rationale Entscheidungsbildung nach dem Prinzip der Gerechtigkeit, aus JP1 und JP2 dagegen die nach jedem beliebigen Prinzip.

Einige Anmerkungen:

1. Das Postulat JP3 braucht nicht akzeptiert zu werden, wenn von einer anderen Gerechtigkeitsauffassung als hier vertreten ausgegangen wird. Auch in diesem Fall garantieren die Postulate JP1 und JP2 die Ableitung der besten Maßnahme entsprechend der jeweiligen Zielvorstellung (von der Gerechtigkeit). Es ist auch denkbar, daß man zwar die hier vertretene Gerechtigkeitsauffassung akzeptiert, bei der konkreten Entscheidungsbildung allerdings nicht von der Zielsetzung Gerechtigkeit ausgeht (also nicht die Gerechtigkeitsannahme macht). In diesem Fall bleibt die Voraussetzung von JP3 nicht erfüllt, JP3 kommt also nicht zur Anwendung; die Entscheidung wird somit auch hier aufgrund der Postulate JP1 und JP2 abgeleitet.

Das Postulat JP3 ist lediglich aus theoretischen Gründen interessant, denn es zeigt, daß im Rahmen einer bestimmten Zielauffassung nur die Kenntnis der Wahrscheinlichkeitswerte für die Sachverhalte genügt, um rational entscheiden zu können. Es ist in der Praxis kaum zu erwarten, daß der Entscheidung die Gerechtigkeitsmatrix, die die Gleichbehandlung auf Kosten der Individualisation garantiert, in ihrer reinen Form zugrunde gelegt wird.

2. Im Unterschied zum konservativen Entscheidungsmodell (vgl. S. 252, wo die Entscheidung völlig vom Sachverhalt determiniert ist (d.h. die Entscheidung wird gewählt, wenn der entsprechende Sachverhalt eine "genügend hohe" Wahrscheinlichkeit erreicht), werden in dem hier aufgestellten Entscheidungsmodell (in welchem nach JP1-JP3 die Entscheidungsbildung auf die Wahrscheinlichkeitsbeurteilung reduzierbar ist) folgende Nachteile vermieden:

Erstens werden hier die Konsequenzen von Entscheidungen berücksichtigt¹⁾ und zweitens läßt sich hier nicht das dort erhaltene unbefriedigende Ergebnis gewinnen, daß wir Freispruch ableiten, obwohl die Wahrscheinlichkeit dafür, daß einer der strafbaren Tatbestände vorliegt, 0.99 beträgt (keiner der Tatbestände jedoch, einzeln betrachtet, eine "genügend hohe" Wahrscheinlichkeit hat). Auf S.252 waren es vier Tatbestände T_1 ("Mord"), T_2 ("Totschlag"), T_3 ("minder schwerer Fall des Totschlags") und T_4 ("Notwehr") mit den Wahrscheinlichkeiten jeweils von 0.33, 0.33, 0.33, 0.01. Nach JP1 - JP3 erhalten wir hier die Entscheidung entsprechend T_1 . D.h. es wird hier die Maßnahme entsprechend dem leichtesten der strafbaren Tatbestände abgeleitet, was angesichts der Wahrscheinlichkeit von 0.99, daß entweder "Mord" oder "Totschlag" oder "minder schwerer Fall des Totschlags" vorliegt, ein befriedigendes Ergebnis darstellt.

Die fundamentale Gerechtigkeitsauffassung erhält somit auch die entscheidungstheoretische Rechtfertigung: Gerech ist eine juristische Entscheidung, wenn auf Grund des mit überwiegender Wahrscheinlichkeit nachgewiesenen Tatbestands eine diesem Tatbestand angemessene (gesetzlich gebotene) Maßnahme gewählt wird. Falls mehr als zwei Tatbestände vorliegen, so wird die Maßnahme gewählt, die dem wahrscheinlicheren Tatbestand aus der durch JP2 eingeschränkten Tatbestandsmenge tatschuldangemessen ist.

1) Wenn nach JP3 die entscheidungstheoretische Beurteilung auf die wahrscheinlichkeitstheoretische reduzierbar ist, so nur deshalb, weil die Nützlichkeitsmatrix GM eine bestimmte, für eine solche Reduktion geeignete Gestalt hat.

3.5 Feststellung der Schuld (der subjektive Tatbestand)

In der Rechtspraxis muß man nicht nur nachweisen, daß der Angeklagte die ihm zur Last gelegte Straftat ("objektiver Tatbestand") begangen hat, sondern auch, daß diese Straftat ihm persönlich im Sinne der Verantwortlichkeit vorzuwerfen ist ("subjektiver Tatbestand"). In diesem Zusammenhang spricht man von der Schuld bzw. Schuldfähigkeit des Angeklagten. Juristisch gesprochen ist Schuld Voraussetzung der Strafbarkeit. Das bedeutet folgendes: Erweist sich, daß der Angeklagte unschuldig bzw. schuldunfähig ist, so wird er freigesprochen (bei Unschuld) oder bei Schuldunfähigkeit in eine Anstalt anstelle des Gefängnisses eingewiesen. Gewöhnlich geht man, wenn nichts dagegen spricht, davon aus, daß der Angeklagte schuldfähig, d.h. sich der Tat und ihrer Folgen voll bewußt, ist, so daß der Nachweis der Schuld/-fähigkeit des Angeklagten in den meisten Fällen nicht durchgeführt wird. In schwerwiegenden Fällen, wie z.B. bei verschiedenen Tötungsverbrechen, ist ein solcher Nachweis jedoch unerläßlich, denn davon hängt es mitunter ab, ob der Angeklagte zu lebenslanger Freiheitsstrafe verurteilt wird oder nicht. Der Nachweis der Schuldfähigkeit des Angeklagten stützt sich dabei primär auf psychologische oder medizinische Sachverständigengutachten. Ein solcher Nachweis ist natürlich sehr problematisch und birgt viel eher die Gefahr eines Irrtums in sich als der des objektiven Tatbestandes.

Bei unserer bisherigen Abhandlung haben wir nur die Wahrscheinlichkeitstheoretische Feststellung des objektiven Tatbestands erörtert. Wie sieht aber ein Beweis aus, wenn man extra den Nachweis der Schuld/-fähigkeit des Angeklagten verlangt? Darauf wollen wir im folgenden etwas näher eingehen.

Man geht zunächst (wie in der herkömmlichen Rechtspraxis) von der Feststellung des objektiven Tatbestands aus. Falls ein solcher nicht mit (kumulativer) überwiegender Wahrscheinlichkeit nachgewiesen ist, erübrigt sich der nachfolgende Beweis der Schuld/-fähigkeit für diesen Tatbestand; der Angeklagte wird dann also nicht nach diesem Tatbestand bestraft. Falls dagegen der objektive Tatbestand mit (kumulativer) überwiegender Wahrscheinlichkeit festgestellt ist, geht man dazu über, die Schuld bzw. Schuldfähigkeit des Angeklagten nachzuweisen. Überwiegt dabei die Schuld bzw. Schuldfähigkeit probabilistisch, so muß der Angeklagte entsprechend dem objektiven Tatbestand bestraft werden, ansonsten wird er freigesprochen (wenn er

unschuldig ist) oder in eine Anstalt eingewiesen (wenn er schuldunfähig ist). Das soll an einem Beispiel gezeigt werden.

Angenommen, der Richter ist der Überzeugung, daß der Angeklagte die Straftat eher begangen als nicht begangen hat; dagegen ist er sich völlig darüber unsicher, ob der Angeklagte sich über die Folgen seiner Tat im klaren war (auch die Sachverständigengutachten helfen ihm dabei wenig). In diesem Fall muß der Angeklagte freigesprochen (oder anders als entsprechend dem objektiven Tatbestand bestraft) werden.

Wir wollen das nachfolgende (ziemlich triviale) Postulat in qualitativer Form formulieren.

Wir benutzen folgende Abkürzungen:

"S" für: "Der Angeklagte ist schuldig bzw. schuldfähig" und

"-S" für: "Der Angeklagte ist unschuldig bzw. schuldunfähig".

">" ist zu lesen als: "wahrscheinlicher"; "≥" als: "nicht weniger wahrscheinlich".

JP4 (Beweis für die Schuld/-Fähigkeit). Wenn $S > -S$, dann wird die Wahl der Maßnahme nach JP1 - JP3 (entsprechend dem objektiven Tatbestand) bestätigt. Wenn $-S \geq S$, dann wird der Angeklagte freigesprochen (im Falle der Unschuld) oder in eine Anstalt eingewiesen (im Falle der Schuldunfähigkeit).

Das obige Postulat läßt sich auch leicht in quantitativer Form angeben. Wir müssen dann nur $S > -S$ bzw. $-S \geq S$ durch $P(S) > P(-S)$ bzw. $P(-S) \geq P(S)$ ersetzen.

Dafür, dieses Postulat eher in qualitativer Form zu gebrauchen, sprechen folgende Gründe:

1. Die beste Entscheidung wird entsprechend dem objektiven Tatbestand abgeleitet. Die Feststellung der Schuld/-fähigkeit dient lediglich der Bestätigung dieser Entscheidung. Also ist der Wahrscheinlichkeitswert für die Schuld/-fähigkeit irrelevant für die Wahl der Maßnahme nach dem objektiven Tatbestand.
2. Beim Nachweis der Schuld/-fähigkeit haben wir es mit nur zwei möglichen Zuständen zu tun: Entweder ist der Angeklagte schuldig/schuldfähig oder der Angeklagte ist unschuldig/schuldunfähig. Es genügt also, wenn die Wahrscheinlichkeit für "Schuld"/"Schuldfähigkeit" überwiegt, denn das bedeutet im Fall, daß nur zwei mögliche Zustände vorliegen, nichts anderes, als daß $P(S) > 0.5$ ist.
3. Im Unterschied zum Nachweis des objektiven Tatbestandes, der sich auf empirische Tatsachen stützt, betrifft die Feststellung der

Schuld/-fähigkeit ein psychologisches Problem, das offensichtlich noch schwerer einer Quantifizierung bzw. der Begründung dieser Quantifizierung seitens des Beurteilenden zugänglich ist. (Obwohl wir natürlich auch hier davon sprechen können, daß der Richter X im Grade von der Schuld/-fähigkeit des Angeklagten überzeugt ist, wobei diese p wie bei der Feststellung der subjektiven Wahrscheinlichkeit des objektiven Tatbestands durch verschiedene Metrisierungsverfahren ermittelt werden kann).

Zur Begründung dessen, warum die Schuld/-fähigkeit 'nur' überwiegend und nicht mit einer bestimmten (höheren) Wahrscheinlichkeit nachgewiesen werden soll, dienen die gleichen Argumente, die bereits bei der Rechtfertigung von JP2 (Forderung nach kumulativ überwiegender Wahrscheinlichkeit bei der Feststellung des objektiven Tatbestandes) angeführt wurden, u.a.:

- 1) Das Kriterium "mit überwiegender Wahrscheinlichkeit"¹⁾ läßt sich formal rechtfertigen, die Festlegung eines beliebigen höheren Wahrscheinlichkeitswertes dagegen nicht, denn eine derartige Festlegung bleibt immer willkürlich.
- 2) Es bleibt dem Richter überlassen, wie sein konkretes 'p' aussieht. JP4 besagt nur soviel, daß sein 'p' auf jeden Fall größer sein muß als 0,5, wenn er die Ableitung der besten Maßnahme entsprechend dem objektiven Tatbestand bestätigen will²⁾.

Man kann sich fragen, wozu das Postulat JP4 (dessen Aussage trivial erscheint) gut ist. Der Grund ist folgender: Das Postulat JP4 bezieht in die rationale juristische Entscheidungsbildung auch den subjektiven Sachverhalt ein, denn mit der Wahl der Maßnahme nach dem objektiven Tatbestand ist der Rechtsprozeß noch nicht abgeschlossen. Durch JP1 - JP4 wird somit der ganze juristische Entscheidungsprozeß³⁾ - von der Feststellung des objektiven Tatbestands bis zur Wahl der Maßnahme nach diesem Tatbestand und die Bestätigung dieser Wahl nach dem subjektiven Tatbestand axiomatisch festgelegt.

1) Wie bereits festgestellt, stimmen die beiden Beweiskriterien "mit überwiegender Wahrscheinlichkeit" und "mit kumulativ überwiegender Wahrscheinlichkeit" miteinander überein, wenn nur zwei Tatbestände vorliegen.

2) Für die Entscheidung bzgl. der Maßnahme reicht ein qualitativer Wahrscheinlichkeitsbegriff aus. Bei der Konkretisierung des Strafrahmens müssen wir hingegen außerdem die sog. "Schwere der Schuld" (die nicht nur durch Vorsatz bzw. Fahrlässigkeit, sondern auch durch das Mitverschulden des Opfers bedingt ist) berücksichtigen. Die Quantifizierung der "Schwere der Schuld" (die keine Wahrscheinlichkeit ist) setzt voraus, daß die Schuld zumindest mit überwiegender Wahrscheinlichkeit nachgewiesen ist. Zur "Schwere der Schuld" vgl. S.399 dieser Arbeit; zur Quantifizierung der Schuld siehe S.399, Fn.4.

3) Von der Konkretisierung der Strafe innerhalb des Strafrahmens sei zunächst abgesehen, siehe dazu S.385ff.

3. Kapitel. Kollektive Entscheidungen

In den beiden ersten Kapiteln dieses Teils beschäftigten wir uns mit der rationalen Entscheidungsbildung eines einzelnen Entscheidungssubjekts (eines Richters). Wir gingen dabei davon aus, daß das Entscheidungssubjekt sowohl subjektive Wahrscheinlichkeitswerte für Sachverhalte (d.h. objektive Tatbestände) wie auch subjektive Nützlichkeiten für Konsequenzen kennen muß, um seine Entscheidungen im juristischen Fall zu treffen. (Im Anhang werden wir außerdem auf die Frage eingehen, wie das Entscheidungssubjekt zu solchen Wahrscheinlichkeiten und Nützlichkeiten kommt (vgl. den Abschnitt über die Metrisierungen). U.a. wird gezeigt, daß die Kenntnis der qualitativen Präferenzordnung der Konsequenzen ausreicht, um die quantitativen Nützlichkeiten abzuleiten, und daß es verschiedene Methoden gibt, zu numerischen Wahrscheinlichkeiten zu gelangen.

Im zweiten Kapitel haben wir vier Postulate JP1-JP4 aufgestellt, die zusammen mit den Savage-Postulaten (vgl. Anhang, S. 431ff) den juristischen Entscheidungsprozeß axiomatisch festlegen. Ein rationaler Richter erhält somit auf konsistente und konsequente Weise seine Entscheidung, wenn er sich im Einklang mit diesen Postulaten verhält.

Geht man bei der rationalen juristischen Entscheidungsbildung von der fundamentalen Gerechtigkeitsauffassung aus, deren Ziel in der formalen Wiederherstellung des juristischen Status Quo besteht, so genügt dazu - wie wir gezeigt haben - bereits die Kenntnis der Wahrscheinlichkeitswerte für die in Frage kommenden Tatbestände, denn in einem solchen Fall reduziert sich die entscheidungstheoretische Beurteilung der Situation auf die wahrscheinlichkeitstheoretische Feststellung des Sachverhalts; nach JP2 bzw. JP2* braucht man dabei nur die Wahrscheinlichkeit bzw. qualitative Wahrscheinlichkeit zu kennen, um sich für die (tatschuldangemessene) Maßnahme zu entscheiden.

In diesem Kapitel wollen wir uns mit einem ganz anderen Problem des juristischen Entscheidungsprozesses befassen, nämlich dem der kollektiven Entscheidungsbildung.

Zur Verdeutlichung der Situation:

Gewöhnlich setzt sich ein bundesdeutsches Gericht in Strafsachen aus einer ungeraden Anzahl von Richtern (Berufsrichtern + ehrenamtlichen Richtern), mindestens einem und höchstens fünf, zusammen. Eine Ausnahme bildet das mit vier Richtern besetzte sog. erweiterte Schöffengericht (Amtsgericht), wo zusätzlich zum einfachen Amtsgericht evtl. auf Antrag der Staatsanwaltschaft bei umfangreichen Strafsachen ein zweiter Berufsrichter zur Hauptverhandlung hinzugezogen ist.

Wir beginnen unsere Ausführungen mit einem kurzen informellen Überblick darüber, welche Entscheidungen von welchen Richtern (Gerichten) nach dem deutschen Recht getroffen werden dürfen. Wir beginnen mit dem sog. Amtsgericht.

Das Amtsgericht setzt sich entweder aus einem Berufsrichter oder aus drei Richtern (Schöffengericht: 1 Berufsrichter + 2 ehrenamtliche Richter) oder aus vier Richtern (erweitertes Schöffengericht: 2 Berufsrichter + 2 ehrenamtliche Richter) zusammen. Straftaten (Vergehen), die mit höchstens 6 Monaten Freiheitsstrafe bedroht sind, werden stets vom Einzelrichter entschieden. Das Schöffengericht (wie auch das erweiterte Schöffengericht) entscheidet über die Straftaten, deren Strafmaß drei Jahre Freiheitsstrafe nicht überschreitet. Sie dürfen weder über Unterbringung in einer psychiatrischen bzw. sozialtherapeutischen Anstalt noch über eine solche in Sicherungsverwahrung entscheiden.

Das Landgericht setzt sich entweder aus drei Richtern (kleine Strafkammer: 1 Berufsrichter + 2 ehrenamtliche Richter) oder fünf Richtern (große Strafkammer, auch Schwurgericht genannt: 3 Berufsrichter + 2 ehrenamtliche Richter) zusammen. Die kleine Strafkammer entscheidet über Berufungen gegen Urteile des Einzelrichters beim Amtsgericht und auch über Verbrechen und Vergehen, für die weder das Amtsgericht noch das Oberlandesgericht zuständig sind. Die große Strafkammer entscheidet sowohl über Berufungen gegen Urteile des Schöffengerichts als auch über Verbrechen und Vergehen, die die Zuständigkeit des Amtsgerichts bzw. des Oberlandesgerichts, aber auch die der kleinen Strafkammer übersteigen (wie z.B. Mord, Totschlag, schwere Sexualstraftaten, Aussetzung, Raub, schwere Brandstiftung etc.).

Das Oberlandesgericht setzt sich entweder aus drei Berufsrichtern (Strafsenat/Revisionsinstanz) oder aus fünf Berufsrichtern (Strafsenat/1. Instanz) zusammen. Der Strafsenat ist in Strafsachen hauptsächlich für die Revision gegen die Berufungsurteile des Landgerichts, ferner für die Sprungrevision gegen amtsgerichtliche Urteile sowie für Beschwerden gegen Beschlüsse des Landgerichts zuständig. Das Oberlandesgericht entscheidet ferner in der Besetzung mit fünf Richtern über schwere Fälle von Friedensverrat, Hochverrat, Landesverrat und bei einigen anderen politischen Delikten (z.B. Völkermord, u.U. Staatsschutzdelikte).

Der Bundesgerichtshof schließlich ist der oberste Gerichtshof des Bundes im Bereich der ordentlichen Gerichtsbarkeit. Die Senate entscheiden in der Besetzung von fünf Berufsrichtern in Strafsachen über Beschwerden, i.d.R. mit drei Berufsrichtern über Revisionen gegen Urteile der großen Strafkammer und der Oberlandesgerichte.

Damit haben wir die Zuständigkeit der verschiedenen Gerichte in groben Zügen beschrieben. Nun wollen wir die Frage der Entscheidungsabstimmung erläutern.

Das Gericht entscheidet, sofern das Gesetz es nicht anders bestimmt, mit der absoluten Mehrheit der Stimmen. Im Strafverfahren ist zu einer jeden dem Angeklagten nachteiligen Entscheidung, welche die Schuldfrage, die Bemessung der Strafe, die Anordnung einer Maßregel¹⁾, über die Besserung und Sicherung betrifft, eine 2/3-Mehrheit der Stimmen erforderlich²⁾, über die Strafaussetzung zur Bewährung wird mit einfacher Mehrheit entschieden. Die Stimmen aller Richter haben gleiches Gewicht (eine Ausnahme bildet das mit vier Richtern besetzte erweiterte Schöffengericht: Bei Entscheidungen, die mit einfacher Mehrheit zu treffen sind, gibt die Stimme des Vorsitzenden den Ausschlag³⁾).

Aus der obigen Erläuterung ersieht man, daß zwar über die Schuldfrage eine Entscheidungsvorschrift (nämlich eine 2/3-Mehrheit) vorhanden ist. Dagegen ist im Hinblick auf die Entscheidung bzgl. des objektiven Sachverhalts keine derartige Vorschrift formuliert. Die Schuldfrage setzt jedoch ganz klar eine Entscheidung bzgl. der Feststellung des objektiven Sachverhalts voraus. Ihre Bestimmung ist daher eine noch zu schließende Lücke.

Die rationale Entscheidungsfindung eines einzelnen Richters (die Einzelentscheidung) haben wir ausführlich in den beiden ersten Kapiteln dieses Teils beschrieben. Wie sie bei drei, vier und fünf

1) Bei 5 Richtern sind es somit mindestens 4 und bei 4 Richtern mindestens 3, die eine 2/3-Mehrheit bilden.

2) D.h. bei Entscheidungen mit einfacher Mehrheit werden anstatt vier fünf Stimmen gezählt, wovon zwei auf den Vorsitzenden entfallen.

3) Die obigen Angaben entstammen Creifelds, Rechtswörterbuch, S.13f.

Richtern aussieht, wird unten näher erläutert. Wir werden uns hier von vornherein auf sog. Konfliktsituationen beschränken, d.h. Situationen, in denen Meinungsverschiedenheiten auftauchen, die die Entscheidungsbildung erschweren (In einem konfliktfreien Fall reduzieren sich die Mehrheitsentscheidungen auf Einzelentscheidungen).

Die Meinungsverschiedenheiten können sowohl den Sachverhalt betreffen (d.h. die subjektiven Wahrscheinlichkeiten, die die einzelnen Richter dem Sachverhalt zuordnen, können variieren) wie auch die Wahl der Entscheidung selbst. Wir müssen dabei hauptsächlich drei Quellen von Meinungsverschiedenheiten unterscheiden: erstens, diejenige bzgl. der Sachverhaltswahrscheinlichkeit, zweitens diejenige bzgl. der Zielsetzung und damit der Nutzenzuordnung und drittens, diejenige bzgl. der Schuldfrage.

Zur Behebung der Meinungsverschiedenheiten bzgl. der Sachverhaltswahrscheinlichkeit wurde im I. Teil, S. 231 eine sehr einfache Methode vorgeschlagen, die in der Bildung des Mittelwertes aus den einzelnen individuellen Wahrscheinlichkeitswerten bestand. Fordert man eine 2/3-Mehrheit, so bedeutet dies, daß bei der Bestimmung, ob ein Sachverhalt mit einer gewissen Wahrscheinlichkeit gilt, 2/3 aller Wahrscheinlichkeitsbewertungen berücksichtigt werden. Die Mittelwertbildung wird dann auf diese Wahrscheinlichkeitsbewertungen angewandt. Welche Wahrscheinlichkeiten berücksichtigt werden, entscheiden die Richter selbst. Die 2/3-Mehrheitsentscheidung ist offensichtlich besonders dann wirksam, wenn die meisten Bewertungen annähernd gleich sind, die übrig gebliebenen dagegen extrem abweichen, wie z.B. im folgenden Fall: Es seien $P_A(H)=0.9$, $P_B(H)=0.94$ und $P_C(H)=0.2$ die individuellen Wahrscheinlichkeiten dreier Richter A, B und C für die Hypothese H. Es ist anzunehmen, daß die Richter A und B mehr Gründe haben, um miteinander zu koalieren, denn der Mittelwert ihrer Wahrscheinlichkeiten ($=0.92$) weicht nur geringfügig von ihren individuellen Werten ab. Auf diese Weise - d.h. indem sie eine 2/3-Mehrheit bilden - können sie also die extrem abweichende Meinung des dritten Richters C unterdrücken. Aus demokratischer Sicht würde man allerdings empfehlen, daß jede einzelne Wahrscheinlichkeitsbewertung in die gemeinsame Wahrscheinlichkeitsbewertung des Sachverhalts eingeht, selbst wenn die einzelnen Wahrscheinlichkeiten voneinander extrem abweichen. Berücksichtigt man im obigen Fall die Meinung des dritten Richters C, so erhält man als Mittelwert den Wahrscheinlich-

keitswert 0.68. In diesem Wert drücken sich also auf demokratische Weise recht unterschiedliche Meinungen bzgl. des Wahrscheinlichkeitswertes des Sachverhalts aus.

Meinungsverschiedenheiten bzgl. der Schuldfrage sind auf analoge Weise zu beheben. Falls die Schuld des Angeklagten nicht quantifiziert¹⁾, sondern lediglich mit überwiegender qualitativer Wahrscheinlichkeit festgestellt wird, liefert hier die juristische Forderung nach einer 2/3 Mehrheit eine Lösung. Der Vorteil der Quantifizierung (vorausgesetzt, die numerischen Wahrscheinlichkeitswerte repräsentieren wirklich die subjektiven Meinungen²⁾) liegt darin, daß jede Meinung gleichermaßen auf demokratische Weise berücksichtigt wird.

Beispiel:

Nehmen wir an, die Richter A und B sind mit überwiegender Wahrscheinlichkeit überzeugt, daß der Angeklagte schuldig ist, C dagegen, daß er unschuldig ist. Nach dem 2/3-Mehrheitsprinzip wird also die Meinung von A und B angenommen. Nun nehmen wir weiter an, daß die auf ehrliche Weise vorgenommene Quantifizierung der Überzeugungen von A, B und C die jeweiligen Wahrscheinlichkeitswerte 0.51, 0.69 und 0.1 für die Schuld ergab. Berücksichtigt man alle diese Bewertungen gleichermaßen, so erhält man als Mittelwert die Wahrscheinlichkeit 0.43. Da insgesamt die "Unschuld" überwiegt, muß somit die Auffassung von C (auch wenn A und B dagegen sind) akzeptiert werden.

Die Mittelwertbildung ist gegenüber der herkömmlichen Entscheidung für "schuldig"/"unschuldig" nach einer 2/3-Mehrheit eine feinere Methode, die nicht nur jede Meinung berücksichtigt, sondern - fair(!) angewandt - auch die Überzeugungsintensität jedes einzelnen Individuums.

Die Voraussetzung der Fairness und Ehrlichkeit bei der praktischen Anwendung der Methode der Mittelwertbildung ist wichtig, denn wie viele praxisrelevante Methoden kann man sie durch das Manipulieren von Werten zur Durchsetzung der eigenen Interessen mißbrauchen.

1) Wie bereits festgestellt, richtet sich die Ableitung einer Maßnahme (des Strafrahmens) einzig und allein nach dem objektiven Tatbestand. Daher ist seine Quantifizierung unerlässlich, will man rational vorgehen (Ausnahme: man vertritt die Gerechtigkeitsauffassung). Die Lösung der Schuldfrage dient hingegen lediglich der Bestätigung bzw. Zurückweisung dieser Maßnahme. Daher ist es nicht erforderlich, die Schuld probabilistisch zu quantifizieren.

2) Es ist nicht ausgeschlossen, daß man versucht, seinen Wahrscheinlichkeitswert zu manipulieren, um so seine Meinung durchzusetzen.

Meinungsverschiedenheiten hinsichtlich der Zielsetzung (oder der Nutzenzuordnung)

Kehren wir zu unserer entscheidungstheoretischen Ausgangssituation zurück. Wir haben damals die Entscheidungsbildung eines einzigen rationalen Richters verfolgt. Falls der Richter als Einzelrichter (wie beim Amtsgericht) die Entscheidung trifft, wird sich an der endgültigen Entscheidungsbildung nichts ändern. In vielen Fällen werden jedoch mehrere Richter an dieser Entscheidungsbildung beteiligt sein. Hier unterscheiden wir zwei Stadien:

1. Stadium: wie beim Einzelrichter

2. Stadium: die Koordination verschiedener Einzelentscheidungen.

Das zweite Stadium bei den Mehrheitsentscheidungen setzt also da an, wo es beim Einzelrichter nichts mehr zu entscheiden gibt.

Abgesehen vom trivialen Fall, wo alle Richter von vornherein dieselbe Meinung bzgl. der optimalen Maßnahme vertreten (dieser Fall fällt somit in die Domäne der Entscheidungstheorie, wie sie im 1. und 2. Kap. dargestellt wurde), haben die meisten Mehrheitsentscheidungssituationen einen spieltheoretischen Charakter.

In der Spieltheorie unterscheidet man gewöhnlich streng zwischen kompetitiven und nicht-kompetitiven Situationen¹⁾. Die ersteren setzen nicht nur eine unterschiedliche Zielsetzung voraus, sondern auch, daß jeder einzelne Beteiligte versucht, seine eigene Zielsetzung (sein eigenes Interesse) durchzusetzen. Der Gewinn eines einzigen bedeutet hier den Verlust aller anderen. Ein typisches Beispiel dafür ist das Schachspiel. Solche kompetitiven Situationen sind auch in der Rechtspraxis möglich: z.B. das 'Spiel' zwischen Anklage und Verteidigung im Straf- oder das von Parteien im Zivilprozeß²⁾.

In nicht-kompetitiven Entscheidungssituationen muß dagegen eine gemeinsame Entscheidung von mehreren bzw. von allen Entscheidenden angestrebt werden, auch wenn sie von unterschiedlichen Zielvorstellungen ausgehen. Die Entscheidenden müssen sich also in einer solchen Situation im Unterschied zu einem kompetitiven Spiel kooperativ verhalten. In der juristischen Praxis werden hauptsächlich die Mehrheitsentscheidungen unter diesen Aspekt der Spieltheorie fallen: Denn unter dem Entscheidungszwang müssen die Richter trotz

1) auch wenn verschiedene Zwischenformen aus beiden möglich sind.

2) Bei dem letzteren sind durch verschiedene Einigungsverfahren auch gemischte, d.h. zum Teil kompetitive und zum Teil nicht-kompetitive aber auch reine, nicht-kompetitive Situationen möglich.

ihrer unterschiedlichen Zielsetzungen eine gemeinsame Entscheidung treffen¹⁾. Diese Art von Entscheidungssituation verlangt also von den am Prozeß beteiligten Richtern Kompromißbereitschaft. Im folgenden werden wir uns mit diesen (d.h. den nicht-kompetitiven kooperativen Mehrheits-) Entscheidungen befassen. (Wir wollen hier allerdings keine Theorie der Mehrheitsentscheidung entwickeln, sondern nur verschiedene Lösungsansätze für Meinungsverschiedenheiten in bezug auf die Zielsetzung diskutieren, dies nicht zuletzt deshalb, weil mir bislang kein spieltheoretischer Ansatz bekannt ist, der den Eigenarten des juristischen Kooperationsbegriffs gerecht wird. Diese Eigenarten sind: 1) eine 2/3-Mehrheit bei gleicher Gewichtung für alle dem Angeklagten nachteiligen Entscheidungen, ansonsten eine Entscheidung zugunsten des Angeklagten; 2) die Tatsache, daß die Kooperation die gemeinsame Zielsetzung, die durch eine entsprechende Nützlichkeitsmatrix repräsentiert ist, betrifft, nicht dagegen eine gemeinsame Auszahlungsfunktion (wie das bei spieltheoretischen Kooperationen gewöhnlich der Fall ist). Die Situation ist also insofern viel einfacher als bei gewöhnlichen kooperativen Spielen, als wir es hier nicht mit "charakteristischen Funktionen", "stabilen Mengen", "Auszahlungsvektoren", "Shapley-Werten" und dergleichen, d.h. dem üblichen Rüstzeug einer kooperativen Spielsituation zu tun haben, sondern lediglich mit einer gemeinsamen Nützlichkeitsmatrix, die wenigstens von einer 2/3-Mehrheit aller Entscheidenden gebildet werden muß. Die Lösung ist daher ziemlich einfach: Kommt keine 2/3-Mehrheit zustande, so wird zugunsten des Angeklagten entschieden, kommt jedoch eine solche zustande, so müssen alle Beteiligten an der Koalition sehen, wie sie bei gleicher Gewichtung ihrer Stimmen sowohl zu einem gemeinsamen Wahrscheinlichkeitswert für den Sachverhalt als auch zu einer gemeinsamen Nützlichkeitsmatrix gelangen. Aber auch in einer solchen, ziemlich einfachen kooperativen Situation können sich Probleme ergeben, etwa bei der Bildung von Koalitionen innerhalb einer Koalition bzw. dem Wechsel von einer zu einer anderen Koalition, der Wahl der günstigsten Koalition und vielleicht noch andere Probleme, auf die wir hier leider nicht eingehen werden.)

1) Man könnte versuchen, auch die Mehrheitsentscheidungen unter einem kompetitiven Gesichtspunkt zu sehen. Danach würde jeder einzelne Richter im Richtergremium bestrebt sein, sein eigenes Interesse gegen die Interessen der anderen Beteiligten durchzusetzen. Dies stünde allerdings im Widerspruch zur Gesetzespraxis, wo eine gemeinsame Lösung des Falls (die eine Einigung mindestens von 2/3 aller Beteiligten voraussetzt) angestrebt wird, auch wenn dabei einzelne Interessen auf der Strecke bleiben.

Zuerst betrachten wir den Fall, wo alle Richter dieselben Maßnahmen in Betracht ziehen.

Wir nehmen an, daß der Richter A bei seiner Zielvorstellung (Zweckvorstellung bzgl. der Strafe) von der Generalprävention (vgl. S.334 dieser Arbeit), der Richter B von einer Resozialisationsauffassung (vgl. S.335f dieser Arbeit) und der Richter C von der Gerechtigkeitsauffassung, wie sie in dieser Arbeit entwickelt wurde, ausgehen. Die beiden ersten Nützlichkeitsmatrizen mögen wie folgt aussehen:

$$\begin{array}{ll} \text{(A) Generalprävention} & \begin{bmatrix} 1 & 0.1 \\ 0 & 0.6 \end{bmatrix} \\ \text{(B) Resozialisierung} & \begin{bmatrix} 0.5 & 0 \\ 0.4 & 1 \end{bmatrix} \end{array}$$

Die Gerechtigkeitsmatrix hat hingegen nach dem bisherigen eine bestimmte Gestalt:

$$\text{(C) Gerechtigkeit} \quad \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

1. Lösungsvorschlag

Die drei Richter sind zum Kompromiß bzgl. der Entscheidung bereit. Jeder von ihnen will jedoch, daß seine Zielsetzung in die gemeinsame Entscheidung eingeht, wobei jeder auch bereit ist, die Zielvorstellungen der anderen zu berücksichtigen.

In diesem Fall bildet man die gemeinsame Nützlichkeitsmatrix, indem man die drei Nützlichkeitswerte für jede Konsequenz zueinander addiert (dadurch werden die subjektiven Meinungen von A, B und C bzgl. der Bewertung der Konsequenzen gleichermaßen berücksichtigt), d.h.

$$\begin{bmatrix} 1+0.5+1 & 0.1+0+0 \\ 0+0.4+0 & 0.6+1+1 \end{bmatrix} = \begin{bmatrix} 2.5 & 0.1 \\ 0.4 & 2.6 \end{bmatrix}$$

bzw. in transformierter Form:

$$\begin{bmatrix} 0.96 & 0 \\ 0.12 & 1 \end{bmatrix}$$

Legt man diese Nützlichkeitsmatrix der gemeinsamen Entscheidungsbildung zugrunde, so erhält man mit JP1 und JP2 "Strafe" für $p \in (0.54, 1]$ und "keine Strafe" für $p \in [0, 0.54]$.

2. Lösungsvorschlag

Die drei Richter sind zwar zu einem Kompromiß bzgl. der gemeinsamen Entscheidung bereit, wollen jedoch die Zielsetzungen der anderen nicht ohne weiteres akzeptieren.

Hier besteht die Lösung des Konflikts in der Bildung einer Koalition. Wir wollen zunächst den Koalitionsbegriff anhand des mathematischen Mengenbegriffs verdeutlichen. Eine Menge aus n Elementen hat 2^n Teilmengen. Eine dreielementige Menge hat somit $2^3 = 8$ Teilmengen. Mit A, B und C als Elementen sind das $\{A\}, \{B\}, \{C\}, \{A, B\}, \{A, C\}, \{B, C\}, \{A, B, C\}, \emptyset$. Die Nullmenge $\emptyset$ (bzw. das unmögliche Ereignis) wird aus mengentheoretischen Gründen zugelassen. Fassen wir diese Mengen je nach Anzahl ihrer Elemente wiederum in einer Menge zusammen, z.B. die einelementigen Mengen in einer Menge, die zweielementigen Mengen in einer anderen usw., so erhalten wir vier Mengen: $\{\emptyset\}; \{\{A\}, \{B\}, \{C\}\}; \{\{A, B\}, \{A, C\}, \{B, C\}\}; \{\{A, B, C\}\}$. Die Menge, die die Nullmenge enthält, braucht uns hier nicht weiter zu kümmern. Die drei anderen veranschaulichen jeweils eine bestimmte Entscheidungssituation. Die Menge $\{\{A\}, \{B\}, \{C\}\}$ veranschaulicht die Situation, in der jeder gegen jeden ist, ohne zu einem Kompromiß bereit zu sein. In einer solchen Situation ist eine gemeinsame Entscheidung nicht möglich. Die Menge $\{\{A, B, C\}\}$ veranschaulicht die Situation, wie sie im 1. Lösungsvorschlag erläutert wurde: A, B und C koalieren miteinander, d.h. sie sind zu gegenseitigen Konzessionen bereit. Die Menge $\{\{A, B\}, \{A, C\}, \{B, C\}\}$ veranschaulicht dagegen eine echte Koalitionsbildung innerhalb einer dreielementigen Menge. Es gibt drei Möglichkeiten, eine derartige Koalition zu bilden: Erstens, wenn A und B gegen C, zweitens wenn A und C gegen B und drittens, wenn B und C gegen A koalieren. Falls zwei gegen einen koalieren, so müssen sie bereit sein, sich wenigstens innerhalb ihrer Koalition auf eine gemeinsame Lösung zu einigen. Da die deutsche Gesetzgebung eine 2/3-Mehrheit in Strafsachen verlangt, wird dann diese Koalitionsentscheidung die endgültige Entscheidung sein. Das soll an einigen Beispielen veranschaulicht werden:

1. Bsp.: Wenn zwei Anhänger der Vergeltungsauffassung gegen einen Gerechtigkeitsanhänger koalieren, wird sich die Entscheidung nach dem Gesichtspunkt der Vergeltung durchsetzen.

2. Bsp.: Wenn ein Anhänger der Vergeltung und ein solcher der Gerechtigkeit gegen einen Anhänger der generalpräventiven Auffassung koalieren, wird sich die Entscheidung nach dem Gesichtspunkt der "mil-

deren" Vergeltung durchsetzen.

3.Bsp. (Extremfall): Wenn ein Anhänger der Vergeltung, der von der generalpräventiven Wirkung der Strafe ausgeht, zusammen mit einem Anhänger der Resozialisierung (aus irgendwelchen Gründen) gegen einen Gerechtigkeitsanhänger koalitiert, so wird sich die Entscheidung nach dem Gesichtspunkt der Vergeltung bzw. der Resozialisierung oder aber auch dem der Gerechtigkeit durchsetzen, je nachdem, von welchen Nützlichkeitsmatrizen die beiden Koalitionspartner ausgehen.

Eine vierelementige Menge hat $2^4 = 16$ Teilmengen: $\emptyset$, $\{A\}$, $\{B\}$, $\{C\}$, $\{D\}$, $\{A,B\}$, $\{A,C\}$, $\{A,D\}$, $\{B,C\}$, $\{B,D\}$, $\{C,D\}$, $\{A,B,C\}$, $\{A,B,D\}$, $\{A,C,D\}$, $\{B,C,D\}$, $\{A,B,C,D\}$. Nur die Mengen $\{A,B,C\}$, $\{A,B,D\}$, $\{A,C,D\}$, $\{B,C,D\}$ weisen die von der Gesetzgebung verlangte 2/3-Mehrheit auf. (Die Menge $\{A,B,C,D\}$ repräsentiert dagegen eine Entscheidungssituation, wie sie im 1.Lösungsvorschlag dargestellt wurde, d.h. eine Situation, in der jeder mit jedem koalitiert). Bei der einfachen Mehrheit können aber auch die zweielementigen Mengen den Ausschlag geben, falls sich darunter der Vorsitzende befindet (auf den Vorsitzenden entfallen bei Entscheidungen, die mit einfacher Mehrheit erfolgen, zwei Stimmen). Natürlich können sich innerhalb der Koalition weitere Koalitionen bilden. Die Entscheidungsbildung verläuft dann analog einer solchen, wie sie oben für das Drei-Richter-Gremium geschildert wurde.

Eine fünfelementige Menge hat $2^5 = 32$ Teilmengen: $\emptyset$, $\{A\}$, $\{B\}$, $\{C\}$, $\{D\}$, $\{E\}$, $\{A,B\}$, $\{A,C\}$, $\{A,D\}$, $\{A,E\}$, $\{B,C\}$, $\{B,D\}$, $\{B,E\}$, $\{C,D\}$, $\{C,E\}$, $\{D,E\}$, $\{A,B,C\}$, $\{A,B,D\}$, $\{A,B,E\}$, $\{A,C,D\}$, $\{A,C,E\}$, $\{A,D,E\}$, $\{B,C,D\}$, $\{B,C,E\}$, $\{B,D,E\}$, $\{C,D,E\}$, $\{A,B,C,D\}$, $\{A,B,C,E\}$, $\{A,C,D,E\}$, $\{A,B,D,E\}$, $\{B,C,D,E\}$, $\{A,B,C,D,E\}$. Nur die Mengen $\{A,B,C,D\}$, $\{A,B,C,E\}$, $\{A,B,D,E\}$, $\{B,C,D,E\}$ haben eine absolute 2/3-Mehrheit¹⁾. Die Entscheidungsbildung innerhalb der Koalition verläuft analog den oben geschilderten Fällen mit drei- bzw. vierelementigen Mengen.

Zusammenfassend läßt sich sagen, daß dieser Lösungsvorschlag Entscheidungssituationen betrifft, in denen nicht jede Zielvorstellung berücksichtigt werden soll. Die Lösung besteht darin, daß man eine Koalition bildet: Bei drei Richtern sind es zwei, bei vier Richtern drei und bei fünf Richtern vier Personen, die mindestens notwendig sind zur Bildung einer 2/3-Koalitionsmehrheit. Die Entscheidungsbil-

1) "mindestens 2/3" von 5 ist $3\frac{1}{3}$. Die kleinste ganze Zahl, die dieser Bedingung genügt, ist somit 4.

dung erfolgt, indem man entweder gleich den 1.Lösungsvorschlag auf die (unterschiedlichen) Zielsetzungen innerhalb der Koalition anwendet, oder, falls Koalitionen innerhalb der Koalition entstehen, wiederum eine 2/3-Mehrheit bildet und erst dann den 1.Lösungsvorschlag anwendet (Natürlich kann ein aus einer engeren Koalition Ausgeschlossener die größere wechseln. Dieses und andere Probleme im Umkreis des Koalitionsbegriffs wollen wir hier jedoch nicht weiter verfolgen).

Nun betrachten wir die Situation, wo unterschiedliche Maßnahmen von verschiedenen Richtern in Betracht gezogen werden. Das kann seinen Grund darin haben, daß die Richter von verschiedenen Tatbeständen (und dementsprechend von verschiedenen Rechtsfolgen) ausgehen, oder auch darin, daß sie zwar dieselben Tatbestände beurteilen, jedoch in bezug auf die Maßnahmen uneins sind (So kann ein Richter als mögliche Rechtsfolgen Gefängnisstrafe oder Geldstrafe (bzw. Gefängnisstrafe & Sicherungsverwahrung etc.), ein anderer dagegen nur Gefängnisstrafe ansetzen). Wir wollen diese Fälle wiederum anhand von Beispielen veranschaulichen.

1.Bsp.: Angenommen, es ist festgestellt, daß X Y getötet hat. Der Richter A ist überzeugt, daß entweder "Mord" oder "fahrlässige Tötung" als mögliche (qualifizierte) Tatbestände und dementsprechend "lebenslänglich" bzw. "eine Freiheitsstrafe bis zu fünf Jahren" als mögliche Maßnahmen in Frage kommen. B und C sind dagegen der Meinung, daß auch der Tatbestand "Notwehr" nicht ausgeschlossen ist.

Es seien folgende (Anfangs-) Wahrscheinlichkeitsverteilungen gegeben:

$$\begin{aligned} P_A(\text{Mord}) &= 0.4, & P_A(\text{fahrlässige Tötung}) &= 0.6, & P_A(\text{Notwehr}) &= 0^1; \\ P_B(\text{Mord}) &= 0.1, & P_B(\text{fahrlässige Tötung}) &= 0.1, & P_B(\text{Notwehr}) &= 0.8; \\ P_C(\text{Mord}) &= 0.1, & P_C(\text{fahrlässige Tötung}) &= 0.2, & P_C(\text{Notwehr}) &= 0.7. \end{aligned}$$

Die Lösung erfolgt durch Mittelwertbildung.

Wir erhalten:

1) Wir lassen hier die Voraussetzung fallen, daß das Subjekt X keinem der Tatbestände den Wahrscheinlichkeitswert 0 zuordnen darf (vgl. S.352). Es sei allerdings betont, daß dies nur dann zulässig ist, wenn andere Personen dem Tatbestand (den X für "praktisch unmöglich" hält) einen Wahrscheinlichkeitswert größer als 0 zuordnen. Um noch einmal daran zu erinnern: Die Voraussetzung auf S.352 soll verhindern, daß eine Maßnahme gewählt wird, die einem "praktisch unmöglichen" Ereignis korrespondiert.

$P(\text{Mord}) = 0.2$, $P(\text{fahrlässige Tötung}) = 0.3$, $P(\text{Notwehr}) = 0.5$.

Somit wird die Tatbestandsmenge nach JP2 auf {Notwehr} eingeschränkt und "Freispruch" als Maßnahme gewählt.

Wären dagegen die folgenden Wahrscheinlichkeitsverteilungen gegeben:

$P_A(\text{Mord}) = 0.4$, $P_A(\text{fahrlässige Tötung}) = 0.6$, $P_A(\text{Notwehr}) = 0$

$P_B(\text{Mord}) = 0.4$, $P_B(\text{fahrlässige Tötung}) = 0.5$, $P_B(\text{Notwehr}) = 0.1$

$P_C(\text{Mord}) = 0.3$, $P_C(\text{fahrlässige Tötung}) = 0.5$, $P_C(\text{Notwehr}) = 0.2$,

dann hätten wir folgendes Ergebnis erhalten:

$P(\text{Mord}) = 0.37$, $P(\text{fahrlässige Tötung}) = 0.53$, $P(\text{Notwehr}) = 0.1$.

In diesem Fall muß die Tatbestandsmenge nach JP2 auf die Tatbestände "fahrlässige Tötung" und "Notwehr" eingeschränkt werden.

Der nächste Schritt ist die Angabe einer gemeinsamen (eingeschränkten) Nützlichkeitsmatrix. Sei eine solche zur Vereinfachung der Darstellung bei allen drei Richtern die Matrix GM:

A,B,C	fahrl. Tötung (0.53)	Notwehr (0.1)
≤ 5 Jahre	1	0
Freispruch	0	1

Somit folgt die Maßnahme "Freiheitsstrafe bis zu fünf Jahren".

Beachte: Falls A und C im ersten Fall eine Koalition gegen B bilden, können sie mit der Matrix GM als 2/3-Mehrheit die Bestrafung wegen fahrlässiger Tötung erzwingen. Denn hier liegt die folgende Wahrscheinlichkeitsverteilung vor: $P(\text{Mord}) = 0.25$, $P(\text{fahrl. Tötung}) = 0.4$, $P(\text{Notwehr}) = 0.35$; es muß somit nach JP2 und JP3 die Maßnahme "Freiheitsstrafe bis zu fünf Jahren" gewählt werden. (Dagegen ergibt die Koalition von C mit B "Freispruch" als die beste Maßnahme). Da C für sich allein keine solche (in seinen Augen vermutlich unbefriedigende) Entscheidung ableiten kann, ist zu erwarten, daß er eine solche Koalitionsbildung mit A vermeiden würde. Dagegen hat in diesem Fall die Berücksichtigung aller Meinungen keinen Einfluß auf die Wahl der Entscheidung, denn sowohl mit als auch ohne A wird "Freispruch" abgeleitet.

2.Bsp.: Angenommen, A zieht neben der Gefängnisstrafe für den Tatbestand "Diebstahl" auch eine Geldstrafe in Betracht. Die Richter B und C sind dagegen der Meinung, daß eine Geldstrafe völlig ausreichend ist. Die Nützlichkeitsmatrizen mögen so aussehen:

A	Diebstahl	kein Diebstahl
Gefängnis	1	0
Geldstrafe	0.5	0
Freispruch	0	1

B,C	Diebstahl	kein Diebstahl
Gefängnis	0	0
Geldstrafe	1	0
Freispruch	0	1

Wir ergänzen somit die Nützlichkeitsmatrix von B und C um eine Nullzeile. An der Entscheidung von B und C ändert sich dabei nichts. Falls B und C mit A koalieren, wird die gemeinsame Nützlichkeitsmatrix lauten:

A,B,C	1	0
	2.5	0
	0	3

oder transformiert

	1/3	0
	5/6	0
	0	1

Somit 'dominiert' die Geldstrafe über die Gefängnisstrafe; die Nützlichkeitsmatrix kann daher auf eine 2x2-Matrix reduziert werden:

	Diebstahl	kein Diebstahl
Geldstrafe	5/6	0
Freispruch	0	1

Für $p \in (0.55, 1]$ wird nach dieser Nützlichkeitsmatrix "Geldstrafe", für $p \in [0, 0.55]$ "Freispruch" abgeleitet.

Falls B und C nicht mit A koalieren, sieht ihre gemeinsame Nützlichkeitsmatrix so aus:

B,C	Diebstahl	kein Diebstahl
Geldstrafe	1	0
Freispruch	0	1

Für $p \in (0.5, 1]$ und JPl wird also "Geldstrafe" und für $p \in [0, 0.5]$

"Freispruch" gewählt. Die Koalition mit A führt also dazu, daß auch dann "Freispruch" abgeleitet wird, wenn die Wahrscheinlichkeit für den Tatbestand "Diebstahl" überwiegt (jedoch 0.55 nicht übersteigt).

Angenommen, die Nützlichkeitsmatrix von A sehe wie folgt aus:

$$A \quad \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}$$

(d.h. A zieht die Geldstrafe als mögliche Alternative gar nicht in Betracht).

Die gemeinsame Nützlichkeitsmatrix hat dann transformiert die Gestalt

$$A, B, C \quad \begin{bmatrix} 1/3 & 0 \\ 2/3 & 0 \\ 0 & 1 \end{bmatrix}$$

Somit muß nach JP1 für $p \in (0.6, 1]$ "Geldstrafe" gewählt werden.

Abschließend noch einige Bemerkungen:

1. Gewöhnlich wird in einer Konfliktsituation ein Richter für die Strafe, ein anderer dagegen, oder einer für die Gefängnisstrafe, ein anderer für die Geldstrafe usw. sein. Das läuft jedoch auf das hinaus, was die Nützlichkeitsmatrix zusammen mit der Wahrscheinlichkeitsmatrix in expliziter Form angibt: Man ist entweder in bezug auf den Tatbestand uneins oder vertritt in der Einschätzung von Maßnahmen bzw. von Konsequenzen jeweils einen anderen Standpunkt. Man wird in der juristischen Praxis außerdem beobachten, daß auch ohne Einigung bzgl. des objektiven Tatbestandes eine Entscheidung getroffen wird. Solche Entscheidungen sind vom hier vertretenen Standpunkt aus natürlich nicht rational.

2. Eine Uneinigkeit kann natürlich auch die normative Begründung des Tatbestandes betreffen. Mit solchen Problemen befaßt sich die juristische Begriffsanalyse. Diese aus dem Auslegungsbestreben entstandene Richtung gehört noch immer in den dogmatischen Bereich der Jurisprudenz. Jedoch wurden hier mit der Arbeit von Killian¹⁾ neue Tendenzen aufgezeigt, dieses sehr schwierige und mitunter unüber-

1) W. Killian, Juristische Entscheidung und elektronische Datenverarbeitung (1974).

schaubare Gebiet zu rationalisieren. Insofern sich meine Arbeit mit empirischen Verfahren befaßt, und dazu gehören die Feststellung von Tatsachen, der Nachweis des Bestehens des Sachverhalts mit empirischen Mitteln und die rationale Entscheidungsfindung im Strafprozeß, fällt die Begriffsanalyse weg. Dennoch soll hier darauf hingewiesen werden, daß zusammen mit der sprachanalytisch orientierten Tatbestandsbegründung und der rationalen Strafzumessung die "rationale juristische Entscheidungstheorie" den Grundriß dessen bildet, was man unter einer "rationalen Rekonstruktion juristischer Praxis" verstehen könnte ¹⁾.

1) Eine "rationale Rekonstruktion juristischer Praxis" kann ferner neben den drei Disziplinen "rationale juristische Entscheidungstheorie", "rationale Strafzumessung" und "rationale Auslegung" ein viertes Gebiet, nämlich "rationale juristische Ethik" enthalten. Was allerdings das letztere angeht, so wurden bislang noch keine Ansätze in diese Richtung hin unternommen, so daß im gegenwärtigen Zustand es verfrüht ist, von einer rationalen juristischen Ethik anders als im hypothetischen Modus zu sprechen.

4. Kapitel. Rationale Strafzumessung

Ein besonders wichtiger Aspekt der strafrechtlichen Tätigkeit ist die Strafzumessung, d.h. die Bestimmung der konkreten Strafe innerhalb des gesetzlichen Strafrahmens im Einzelfall. Es ist klar, daß eine solche nur dann einigermaßen objektiv sein kann, wenn sie rational hergeleitet ist. Das bedeutet aber, daß man bestimmte Grundsätze ("Strafzumessungsgrundsätze") haben muß, nach denen sich die Konkretisierung der Strafe im Einzelfall richten kann. Im Unterschied zur herkömmlichen Strafzumessungspraxis, die gerade durch das Fehlen solcher Kriterien charakterisiert ist¹⁾ versucht man neuerdings, wenigstens in der Theorie diesem Mangel durch Erwägung mathematischer Optimierungsmethoden bei der Strafzumessung abzuhelpfen. In dieser Hinsicht ist vor allem das Modell von Haag²⁾ von Bedeutung. Wir werden im folgenden uns dieses Modell näher anschauen.

Das Modell von Haag erweist sich allerdings in mancher Hinsicht noch als unbefriedigend:

1. Haag begreift nicht nur die Wahl einer konkreten Strafe (Strafdauer bzw. Geldhöhe), sondern auch die Wahl eines Strafrahmens als eine solche unter Sicherheit. In unserem Modell wird aber die Wahl eines Strafrahmens als eine Entscheidung unter Risiko verstanden. Erst die Bestimmung der Strafdauer bzw. Geldhöhe reduziert sich für uns auf ein Optimierungsproblem unter Sicherheit.
2. Die empirische Bestimmung der Tatschuld bzw. der tatschuldangemessenen Strafe ist von Haag unbefriedigend gelöst, denn er gibt kein Verfahren an für die Konstruktion geeigneter Skalierungsintervalle für die einzelnen Tatschuldkomponenten, welche die Tatschuld ergeben. (Die willkürliche Wahl der Skalierungseinheit führt aber zu unterschiedlichen und sogar unsinnigen Ergebnissen, je nachdem, welche Skalierungseinheit verwendet wird). Es ist sogar zu vermuten, daß ein solches Verfahren zwar bei der globalen Tatschuldbestimmung, jedoch nicht bei einer solchen aus den einzelnen Tatschuldkomponenten angebbar sein kann.
3. Die Haagsche Zielsetzung erweist sich als zu eng und ohne theoretische Fundierung. Sie wird im Sinne des 2.Kap. (dieses Teils) liberalisiert.

Gegenüber dem empirischen Modell von Haag wird hier ein numerisches Modell zur Bestimmung der tatschuldangemessenen Strafe innerhalb des Strafrahmens aufgestellt. Allerdings wird (wie bei Haag) auch hier die Voraussetzung gemacht, daß der funktionale Zusammenhang zwischen einer Strafe und der Tatschuld einen bestimmten (logarith-

1) Vgl. K. Haag, Rationale Strafzumessung (1970), S.15: "Dieser Mangel an Grundsätzen, die eine Bestimmung der Strafe auf Grund rationaler Überlegung ermöglichen würden, führt dazu, daß die Strafzumessung weitgehend gefühlsmäßig-intuitiv vorgenommen wird und daß darin die Eigenarten der Richterpersönlichkeiten, außerrechtliche Motive, insbesondere ideologische Momente usw. sich niederschlagen. Durch die höheren Instanzen ist mangels ausreichender rationaler Kriterien eine genügend wirksame Kontrolle nicht möglich. In der Urteilsbegründung zum Strafmaß herrschen meist stereotype Wendungen vor, die die Verlegenheit widerspiegeln, ein auf andere Weisen gefundenes Strafmaß begründen zu müssen".

2) Haag, ibid.

mischen) Charakter hat. Da das Hauptziel der juristischen Strafzumessung in der Bestimmung der "tatschuldangemessenen" Strafe innerhalb des Strafrahmens liegt, werden wir uns im wesentlichen auch auf diesen Bereich der rationalen Entscheidungsbildung beschränken.

Wir wollen den Leser noch einmal daran erinnern, daß es sich bei der in dieser Arbeit aufgestellten rationalen juristischen Entscheidungstheorie um ein zweistufiges Entscheidungsmodell handelt: Die erste Stufe betrifft die Wahl einer juristischen Maßnahme (d.h. eines Strafrahmens), was eine "Entscheidungswahl unter Risiko" darstellt, die zweite die Konkretisierung des Strafrahmens, d.h. die Wahl einer bestimmten Strafe innerhalb des Strafrahmens, was eine "Entscheidungswahl unter Sicherheit" ist.

Um das Strafzumessungsmodell (Optimierungsmodell) mit dem oben entwickelten entscheidungstheoretischen Modell, in dem die Wahl der Maßnahme (des Strafrahmens) als eine unter Risiko verstanden wurde, in Einklang zu bringen, werden wir bei der Berechnung der "Normal"- und tatschuldangemessenen Strafe innerhalb des Strafrahmens die Wahrscheinlichkeit berücksichtigen, mit der der strafbare Tatbestand festgestellt ist.

Ferner werden die "Schwere der Tat" und die "Schwere der Schuld" (welche mit der Wahrscheinlichkeit für den Tatbestand bzw. für das Verschulden des Angeklagten nichts zu tun haben) berücksichtigt.

Die Berücksichtigung der Wahrscheinlichkeit für den objektiven Tatbestand bei der Konkretisierung der Strafe dient zur Verhinderung dessen, daß unabhängig davon, mit welcher Wahrscheinlichkeit der strafbare Sachverhalt festgestellt ist (etwa mit 0.51 oder 0.99) ein und dieselbe konkrete Strafe gewählt wird. Die Berücksichtigung der Schwere der Tat bzw. der Schwere der Schuld soll verhindern, daß bei einem leichteren Vorgehen (wie etwa bei einem Diebstahl von 20 DM) bzw. bei einem leichteren Verschulden (wenn z.B. Fahrlässigkeit oder Mitverschulden des Opfers vorliegen) die gleiche konkrete Strafe gewählt wird wie bei schwereren Taten (z.B. dem Diebstahl von 10000 DM) bzw. bei einem schweren Verschulden.

Die Berücksichtigung der Wahrscheinlichkeit für den strafbaren Tatbestand, der Schwere der Tat und der Schwere der Schuld findet ihren Niederschlag in der Modifizierung (Verringerung) der gesetzlich vorgegebenen oberen Grenze des entsprechenden Strafrahmens.

4.1. Das rationale Strafzumessungsmodell von Haag

Haag's Ansatzpunkt ist der, daß man bei der Anwendung von Optimierungsmethoden auf die richterliche Entscheidungspraxis von folgendem Problem ausgeht: Unter Berücksichtigung bestimmter Ziele (z.B. Straftatenverhütung) und bestimmter Restriktionen (z.B. Übermaßverbot¹⁾) soll die optimale Entscheidung getroffen werden.

Das Modell von Haag kann durch ein 5-Tupel $\langle S, Z, R, g, w \rangle$ charakterisiert werden, wobei gilt:

- (1) S ist eine nicht leere und endliche Menge von Maßnahmen²⁾. Diese Maßnahmen kürzen wir mit $S_1, \dots, S_n$ ab.
- (2) Z ist eine nicht leere und endliche Menge, deren Elemente Funktionen (Zielfunktionen) von S in $[0,1]$ sind, d.h. Funktionen, die jeder Maßnahme aus S ein sog. "Maß der Wirksamkeit" zur Erreichung des durch eine Funktion repräsentierten Ziels zuordnen. Die Elemente von Z (Ziele) bezeichnen wir mit $Z_1, Z_2, \dots, Z_r$.
- (3) R ist eine nicht leere und endliche Menge, deren Elemente strafrechtliche Restriktionen³⁾ sind. Die Elemente von R bezeichnen wir mit $r_u^{(j)}, r_o^{(j)}, r_t^{(j)}$, $j = 1, \dots, s$ (der obere Index j verweist darauf, hinsichtlich welcher Straftat die Restriktionen gelten; $r_u^{(j)}$ bezeichnet die untere Grenze der Strafe, die auf eine bestimmte Straftat folgt, $r_o^{(j)}$ die obere Grenze dieser Strafe und $r_t^{(j)}$ die der Straftat proportionale bzw. tatschuldangemessene obere Grenze). Es gilt:

$$r_u^{(j)} \leq S_1^{(j)} \leq r_o^{(j)}, \quad \sum_{S_i \in S} S_i^{(j)} \leq r_t^{(j)}.$$

1) Vgl. S. 334, Fn. 1 dieser Arbeit.

2) Haags Maßnahmen sind nicht mit den Maßnahmen aus den ersten beiden Kapiteln dieses Teils identisch. Zum Haagschen Maßnahmenbegriff siehe unten. Im Modell werden nicht Maßnahmen als solche, sondern als auf positive ganze Zahlen abgebildet betrachtet (denn die Strafen werden in vollen Tagen oder in Tagessätzen (bei Geldstrafen) angeordnet). Die Menge S ist beschränkt, denn ihr kleinstes Element ist nicht kleiner als 1 und ihr größtes Element bestimmt nicht größer als 54750 (≈ 150 Jahre, etwa bei lebenslänglich"). Was die Geldstrafe betrifft, so beträgt diese nach §40 StGB mindestens 5 und (von wenigen Ausnahmen abgesehen) höchstens 360 Tagessätze. Ein Tagessatz wird auf mindestens 2 und höchstens 10000 DM festgesetzt. Das Geldintervall ist somit $[10, 36 \cdot 10^5]$. Die Umrechnung einer Gefängnisstrafe in eine Geldstrafe und umgekehrt ist einfach: Einem Tagessatz entspricht ein Tag Freiheitsstrafe. Aus diesem Grund können wir die Menge S auf Freiheitsstrafen beschränken.

3) Auch diese werden nicht als solche, sondern als auf positive ganze Zahlen abgebildet betrachtet.

- (4) g ist eine Funktion ("Gewichtsfunktion") von Z in $[0,1]$;
 g ordnet jedem $Z_k \in Z$ eine reelle Zahl q mit $0 \leq q \leq 1$, sein
 "Gewicht", zu.
- (5) w ist eine Funktion von S in $\mathbb{R}$, für die gilt:
 für alle $S_i \in S$ $w(S_i) = \sum_{k=1} q(Z_k) \cdot Z_k(S_i)$.
 w bezeichnet die "Gesamtwirksamkeit" einer konkreten Maßnahme
 zur Erreichung aller in Frage kommenden Ziele.

Die mit dem Haag-Modell verbundene normative Regel besagt:
 Wird das Entscheidungsobjekt durch das Haag-Modell $\langle S, Z, R, g, w \rangle$
 charakterisiert, so soll es ein $S \in S$ wählen, für das $w(S) \rightarrow \max$.

Erläuterungen:

(1) Haag gibt folgende Maßnahmen als "zulässig" an:

Strafen:

s1 Geldstrafe, s2 Freiheitsstrafe, s3 Strafaussetzung zur Bewährung
 mit Auflagen und Weisungen.

Nebenstrafen:

n1 Aberkennung der Amtsfähigkeit und der Wählbarkeit, n2 Aberkennung
 des Stimmrechts, n3 Fahrverbot, n4 Einziehung, n5 Bekanntmachungs-
 befugnis.

Maßregeln:

m1 Unterbringung in einer Heil- und Pflegeanstalt, m2 Unterbringung
 in einer Trinkerheilanstalt oder Entziehungsanstalt, m3 Sicherungs-
 verwahrung, m4 Untersagung der Berufsausübung, m5 Entziehung der
 Fahrerlaubnis und Sperre, m6 Polizeiaufsicht, m7 Unbrauchbarmachung,
 m8 Einziehung (sofern es sich nicht um eine Nebenstrafe handelt)
 (vgl. Haag, S.53f).

Haags Modell umfaßt beide Aspekte der Wahl einer strafrechtlichen
 Maßnahme: sowohl eine solche des Strafrahmens (durch die Angabe
 der oberen und der unteren Grenze des Strafrahmens) wie auch eine
 solche der (partiellen¹⁾) Quantifizierung des Strafrahmens. Wie
 man allerdings sehen kann, faßt Haag diese beiden Entscheidungs-

1) partiell, weil nur die (durch eine ganze Zahl repräsentierte)
 Grenze, die der tatschuldangemessenen Strafe entspricht, angegeben
 wird, über die die konkrete Strafe nicht hinausgehen darf.

tätigkeiten als solche unter Sicherheit auf. In unserem Modell wird hingegen die Wahl einer Maßnahme im Sinne des Strafrahmens als eine Entscheidung unter Risiko verstanden. Erst die Bestimmung der Straf- bzw. Geldhöhe reduziert sich für uns auf ein Optimierungsproblem unter Sicherheit. Wir werden später aus diesem Grund das Haag-Modell für unsere Zwecke modifizieren.

(2) Eine Zielfunktion $Z_k \in Z$ ordnet jeder Maßnahme $S_i \in S$ eine reelle Zahl aus dem Intervall $[0,1]$, ihr sog. "Maß der Wirksamkeit zur Erreichung dieses Ziels"¹⁾ zu.

Nach Haag sind folgende Ziele zulässig:

1. Verhütung zukünftiger Straftaten und Verbrechensbekämpfung,
2. Genugtuung für den Verletzten (bei Antrags- und Privatklagedelikten),
3. Vermeidung einer übermäßigen Belastung des Täters,
4. Vermeidung sozialen Drittschadens,
5. Vermeidung von Verfahrens- und Vollstreckungskosten.

Das wichtigste Strafziel ist für Haag das erste. Wir werden hier nicht auf Haags Begründung der Zulässigkeit bzw. Plausibilität der oben angeführten Strafziele weiter eingehen (vgl. dazu Haag, S.19-34), wollen allerdings anmerken: Bei der obigen Klassifizierung fällt ins Auge, daß zwar die Interessen der Gesellschaft bzw. des Staates (v.a. in bezug auf Sicherheit und Kostenminimierung und sogar die des Täters voll zur Geltung kommen, dagegen auf die des Opfers bzw. der Angehörigen des Opfers anscheinend wenig Rücksicht (etwa durch den strafrechtlich zulässigen Übelsausgleich seitens des Täters) genommen wird. (Ausgenommen das unter 2. angeführte Strafziel, das allerdings nur bei Privatklage- und Antragsdelikten berücksichtigt wird und außerdem keineswegs durch das "objektive" Urteil des Richters adäquat wiedergegeben werden kann).

Da der Hauptleidtragende bzw. die Hauptleidtragenden der begangenen Tat das Opfer bzw. seine Angehörigen sind, und nicht die Gesellschaft oder der Staat, scheint die Überbetonung der strafrechtlichen Zielsetzung zugunsten der Staatssicherheit etwas fragwürdig. Ferner kann Haags Konzeption bzgl. der strafrechtlichen Zielsetzungen nicht im Sinne einer Theorie des juristischen Nutzens verstanden werden. Da-

1) Haags Bezeichnung für den Nutzen, vgl. Haag, S.206.

zu fehlt es ihr nicht nur an theoretischer Fundierung, sondern v.a. auch an jeglicher Stellungnahme zum Begriff des (juristischen) Nutzens. Die Frage, was in einer konkreten strafprozessualen Entscheidungssituation für alle Betroffenen vom Nutzen sein könnte, bleibt in Haags Modell unbeantwortet.

(3) Für die Maßnahmen gelten folgende Restriktionen:

1. der gesetzliche Strafraum (repräsentiert durch die obere und die untere Grenze des Strafraums),
 2. die Tatschuld als die obere Grenze für die Gesamtheit der in einem konkreten Rechtsfall verhängten Maßnahmen (das Übermaßverbot).
- Jede Maßnahme muß somit mindestens drei Restriktionen gehorchen:

$$s_u \leq s_j \leq s_o$$

s_u, s_o untere bzw. obere Grenze des Strafraums (der Strafe s_j).

$$n_u \leq n_j \leq n_o$$

n_u, n_o zeitliche Begrenzung der Nebenstrafe (außer n_4 und n_5).

$$m_u \leq m_j \leq m_o$$

m_u, m_o zeitliche Begrenzung der Maßregeln (außer m_6, m_7 und m_8).

n_4, n_5, m_7, m_8 können nur zwei Werte annehmen: Verhängung oder Nichtverhängung einer Maßnahme.

Für die Tatschuldrestriktion gilt:

$$\sum_{j=1}^3 d_j s_j + \sum_{j=1}^5 e_j n_j \leq t$$

d_j, e_j bezeichnen das Gewicht (die Schwere) der einzelnen verhängten Strafarten bzw. Nebenstrafen, der auf der linken Seite stehende Ausdruck stellt die Gesamtbelastung durch Strafen und Nebenstrafen dar, während mit "t" die der Tatschuld entsprechende Strafe abgekürzt ist.

$$\sum_{j=1}^8 h_j m_j \leq o$$

h_j bezeichnet die Schwere (Belastung) der entsprechenden Maßregel m_j , o die proportionale obere Grenze.

(vgl. Haag, S.57f).

Im formalen Modell hatten wir s_j (für Strafarten), n_j (für Nebenstrafen) und m_j (für Maßregelarten) zusammenfassend mit S_i (= Maßnahmen) abgekürzt. (Allerdings entsprechen die S_i im Modell eher den Strafen innerhalb aller möglichen Arten der Bestrafung als der Letzteren selbst). Entsprechend werden s_u, n_u, m_u einheitlich

mit $r_u^{(j)}$ und s_o, n_o, m_o einheitlich mit $r_o^{(j)}$ abgekürzt (der obere Index j verweist darauf, hinsichtlich welcher Straftat die Restriktionen gelten); und schließlich kürzten wir t ($\hat{=}$ die tatschuldangemessene Grenze bei Strafen und Nebenstrafen) und o ($\hat{=}$ die proportionale obere Grenze bei Maßregeln) mit $r^{(j)}$ ab. (Ferner wurde auf die Berücksichtigung von Gewichten d_j, e_j und h_j im Modell verzichtet, denn es ist nicht ganz klar, wozu eine solche Gewichtung gut sein soll; außerdem wird die Funktionsweise dieser Operation von Haag nicht näher erläutert).

(4) Weiter gilt nach Haag, daß alle Strafziele im Verhältnis ihres Gewichts anzustreben sind. Die Strafziele sind also nicht gleichwertig. Durch die Annahme einer Funktion g , die jedem Ziel ein Gewicht zuordnet, haben wir in der Bedingung (4) dieser Forderung Rechnung getragen.

(5) w ordnet jeder Maßnahme ihre sog. "Gesamtwirksamkeit" zur Erreichung aller Strafziele im Verhältnis ihres Gewichts zu. Die Gesamtwirksamkeit (Gesamtzielfunktion) muß maximiert (oder minimiert, falls das Ziel in der Vermeidung eines Zustandes besteht) werden.

Definitionen:

Eine zweckrational richtige Maßnahme ist diejenige, mittels derer alle bei der Strafzumessung zu beachtenden Ziele im Verhältnis ihres Gewichts am weitgehendsten erreicht werden (Haag, S.39).

Eine wertrational zulässige Maßnahme ist diejenige, die sich in den Grenzen der (rechtlich vorgeschriebenen) wertrationalen Restriktionen hält (Tatschuld¹), proportionale Obergrenze). (Haag, S.39f).

Eine sowohl zweckrational richtige als auch wertrational zulässige Maßnahme ist nach Haag eine optimale Maßnahme.

1) "Tatschuld" umfaßt bei Haag die "Schwere des Unrechts" der begangenen Tat (der objektive Tatbestand) und die "Schwere des Verschuldens" (Vorsatz und Fahrlässigkeit) des Täters (der subjektive Tatbestand); vgl. Haag, S.35: "Tatschuld" als Maß der Restriktion der Strafe beinhaltet nicht nur Schuld im dogmatischen Sinn des Strafbegriffs, als Vorsatz und Fahrlässigkeit, Fehlen von Schuld-ausschließungsgründen und Zurechnungsfähigkeit, sondern vor allem auch die "Schwere der Tat", genauer die Schwere des Unrechts der begangenen Tat". Die Strafbegrenzung nach dem Tatschuldprinzip orientiert sich an der begangenen Tat, d.h. die Bemessung der Strafe ist stets unrechtsbezogen (Unrecht $\hat{=}$ objektiver Sachverhalt).

4.2. Die Bestimmung der Tatschuld (bzw. der tatschuldangemessenen Strafe) als Messung

Haag stellt in Anlehnung an das Fechnerische Gesetz die folgende Hypothese auf: Mit zunehmender absoluter Strafhöhe nimmt die einer gleichbleibenden Strafdifferenz entsprechende Tatschulddifferenz (bzw. die vom Täter wahrgenommene Belastung durch die Strafe) ab.

Bezeichnet S die Höhe der Strafe, R das Ausmaß des empfundenen Übel ($\hat{=}$ Belastung des Täters durch die Strafe), so lautet das Fechnerische Gesetz, übertragen auf die Rechtspraxis:

$$R = k \cdot \log \frac{S}{S_a}$$

k ist eine (vom Individuum abhängige) Proportionalitätskonstante, S_a (absolute Reizschwelle), juristisch: die geringste noch mögliche Strafbelastung ($\hat{=}$ die absolute Bestrafungsgrenze im Strafraumen), definiert als $S = S_a$ für $R = 0$.

(Dieser Zusammenhang wird als "Belastungsfunktion" $R = f(S)$ bezeichnet).

Das Tatschuldüberschreitungsverbot (Übermaßverbot) wird somit folgendermaßen interpretiert: Das vom Täter beim Vollzug der auferlegten Strafe empfundene Übel¹⁾ darf die durch das Maß der Tatschuld gesetzte Grenze nicht überschreiten (vgl. Haag, S.65), und gleichen Unterschieden im Tatschuldgehalt sollen gleiche Unterschiede in der Empfindung der Strafe entsprechen.

Somit bestehen nach Haag folgende Beziehungen: Zwischen dem Tatschuldgehalt und der empfundenen Strafbelastung eine lineare und zwischen dieser und der Art und Höhe der Strafe eine nichtlineare (etwa logarithmische) Beziehung.

1) Bezüglich des Ausdrucks "das vom Täter empfundene Übel" vgl. Haag, S.66: "Praktisch maßgebend ist ja nicht die Belastung, die der Bestrafte tatsächlich empfindet, sondern maßgebend ist, was nach der 'Empfindung des Richters' der Bestrafte als Strafübel wahrnimmt. Denn der Richter hat in der Regel keinen Kontakt zu dem Bestraften und er hat die Strafbelastung auch nicht selbst erlebt. In der Regel kennt der Richter nicht einmal die Wirklichkeit des heutigen Strafvollzugs. Er kann die Belastung, die Empfindung des Strafübels nur durch hypothetische "Einfühlung" in den Bestraften ermitteln. Darin steckt die Möglichkeit, daß die tatsächliche Belastungsfunktion völlig verfehlt wird".

Bei der Bewertung der Tatschuld geht man folgendermaßen vor:

1. Die Strafraumen stellen die ersten Bewertungsrahmen der Tatschuld dar (Haag, S.72).
2. Weiter bestimmt man den Durchschnittsfall bzw. die "Normalstrafe" (Haag versteht unter dem "Durchschnittsfall" den Fall, dessen Tatschuldgehalt das arithmetische Mittel derjenigen Tatschuldgehalte darstellt, die der oberen und der unteren Strafraumengrenze entsprechen (Haag, S.73)). Da zwischen Tatschuld und Strafe kein linearer Zusammenhang besteht (s.o.), liegt die dem Durchschnittsfall entsprechende Strafe nicht in der arithmetischen Mitte des Strafraumens. Die Durchschnittsstrafe ist (nach dem Fechnerischen Gesetz) so zu bestimmen:

S_1, S_2 unterer bzw. oberer Strafraum, t_1, t_2 entsprechende Tatschuld

$$t_1 = k \cdot \log \frac{S_1}{S_a} \qquad t_2 = k \cdot \log \frac{S_2}{S_a}$$

$$\begin{aligned} t_m &= \frac{1}{2} \cdot (t_1 + t_2) = \frac{1}{2} \cdot k \cdot (\log \frac{S_1}{S_a} + \log \frac{S_2}{S_a}) \\ &= \frac{1}{2} \cdot k \cdot \log \frac{S_1 \cdot S_2}{S_a^2} = k \cdot \log \frac{\sqrt{S_1 \cdot S_2}}{S_a} \end{aligned}$$

$$\text{Aus } t_m = k \cdot \log \frac{S_m}{S_a} = k \cdot \log \frac{\sqrt{S_1 \cdot S_2}}{S_a} \quad \text{folgt: } S_m = \sqrt{S_1 \cdot S_2}.$$

t_m ist das arithmetische Mittel der Tatschuld und S_m die entsprechende Strafe ("Normalstrafe"). Die dem "Durchschnittsfall" entsprechende "Normalstrafe" ist somit das geometrische Mittel der unteren und oberen Strafraumengrenze (vgl. Haag, S.74).

3. Berücksichtigung anderer Erwägungen (etwa spezial- oder generalpräventiver Art). Eine bestimmte Strafe kann dabei wie folgt zusammengesetzt sein:

$$S = S_m + sp + gp + r$$

S Strafe für das Delikt

S_m Strafe, die der Mitte des Tatschuldrahmens des Delikts entspricht.

sp durch spezialpräventive Erwägungen bewirkte Strafänderung

gp durch generalpräventive Erwägungen bewirkte Strafänderung

r durch sonstige Faktoren bewirkte Strafänderung

(damit $S = S_m$ ist, d.h. damit die Strafe S der mittleren tatschuldangemessenen Strafe entspricht, müssen sich die Komponenten s_p , g_p und r gegenseitig aufheben).

4. Empirische Bestimmung der tatschuldangemessenen Strafe. Nach Haag wird die Tatschuld wie folgt bestimmt (S.109):

"Alle strafbegründenden Merkmale sind zuerst multiplikativ zu verbinden (Produkt P), alle strafferhöhenden oder strafmildernden Umstände sind mit dem vorigen Produkt zu multiplizieren und dann zu P zu addieren. Liegt ein strafbegründendes Merkmal nicht vor, so hat es den Skalenwert 0, so daß das ganze Produkt und damit auch die Tatschuld $t = 0$ ist. Dagegen bewirkt das Nichtvorliegen eines strafferhöhenden oder -mildernden Umstandes nur, daß sich das Produkt P und damit die Tatschuld t nicht verändert.

$$t = P \cdot (1+A) = P + PA = b_1 \cdot b_2 \cdot \dots \cdot b_k + P \cdot (a_1 + a_2 + \dots + a_l)$$

$b_1, \dots, b_k$ strafbegründende Merkmale, Produkt P

$a_1, \dots, a_l$ strafändernde Merkmale, Summe A

Die Gewichtung der einzelnen Merkmale erfolgt durch Wahl einer entsprechenden Skaleneinheit.

Die Anwendung dieser Regel auf die allen Straftaten gemeinsamen Komponenten ergibt:

$$t = x_1 \cdot x_2 \cdot x_3 \cdot x_4 (1 + a_1 + a_2 + \dots)$$

x_1, x_2 Unrechtssachverhalt¹⁾ Unrechtsvorwurf²⁾

x_3, x_4 Schuldssachverhalt³⁾ Schuldvorwurf⁴⁾

a_1, a_2 strafändernde Merkmale."

(die tatschuldangemessene Strafe S_t ergibt sich dann aus der Formel

$$t = k \cdot \log \frac{S_t}{S_a}).$$

Will man Haags Methode zur empirischen Bestimmung der Tatschuld und der tatschuldangemessenen Strafe (aus der so festgestellten Tatschuld) praktisch anwenden, so muß man geeignete Skalierungsintervalle für die Tatschuld konstruieren. Diese Skalierungsintervalle müssen je nach Strafraum verschieden ausfallen. Denn nur so erhalten wir das einer konkreten Straftat bzw. Straftat adäquate Tatschuldmaß⁵⁾.

1) Schwere des Unrechts (Rechtsgutverletzung), 2) soziale Verwerflichkeit der Tat, 3) Schwere der Schuld, 4) individuelle Verwerflichkeit. 5) Haags Ansatz ist unbefriedigend, da er auf die Konstruktion von Skalierungsintervallen nicht eingeht. Die willkürliche Wahl von solchen bzw. der Skalierungseinheit führt aber zu unterschiedlichen und sogar unsinnigen Ergebnissen, je nachdem, welche Skalierungsintervalle bzw. Skalierungseinheiten verwendet werden. Bei der obigen Dar

Bei der empirischen Tatschuldbestimmung müssen wir somit als erstes die Tatschuldgrenzen angeben, die - legt man der Strafzumessung die obige Hypothese über den logarithmischen Zusammenhang zwischen Strafen und Tatschuld zugrunde - logarithmisch auf die untere und die obere Grenzen des Strafrahmens abgebildet werden müssen. Sodann setzen wir die Tatschuldgrenzen den Grenzen des Skalierungsintervalls gleich. Das soll im folgenden an einem konkreten Beispiel näher gezeigt werden.

Seien folgende Daten gegeben:

$S_1 = 30$ Tage, untere Grenze des Strafrahmens

$S_2 = 5$ Jahre bzw. 1827 Tage, obere Grenze des Strafrahmens.

Diesen Grenzen entsprechen 0 und 4.1 als die untere und obere Grenze des Tatschuldrahmens bzw. Skalierungsintervalls.

Genau genommen haben wir hier stillschweigend die Voraussetzung gemacht, daß $k = 1$ und S_a (die geringste noch wahrnehmbare Belastung innerhalb des Strafrahmens) = S_1 (die untere Grenze des Strafrahmens) sind. Wir erhalten somit:

$$t_1 = 1 \cdot \log 1 = 0 \text{ und } t_2 = 1 \cdot \log 60.9 \approx 4.1$$

(t_1, t_2 sind untere bzw. obere Grenze des Tatschuldrahmens).

Da k und S_a nach der Haagschen Methode von Fall zu Fall und von Individuum zu Individuum variieren, müssen sie in einem konkreten Fall vor der Skalierung empirisch (individuell) ermittelt werden. Es ist klar, daß sich unterschiedliche Skalierungsintervalle ergeben, je nachdem, wie k und S_a aussehen.

Als nächstes wird empirisch (durch verschiedene Skalierungsverfahren, wie Rating, Intervall-, Verhältnisbestimmung und dergl.¹⁾) festgestellt, wo in diesem Intervall die Tatschuld für die aktuelle Straftat liegt. Liegt eine solche etwa in der Mitte des Skalierungsrahmens (z.B. bei 2.05), so erhalten wir die Strafe von 234 Tagen. (Die tatschuldangemessene Strafe entspricht in diesem konkreten Fall somit der Normalstrafe S_m . Denn mit der auf S.392f angegebenen logarithmischen Methode errechnen wir: $S_m = \sqrt{30 \cdot 1827} \approx 234$ Tage). Wird $t = 1$ angegeben, so erhalten wir die tatschuldangemessene Strafe von 81 Tagen.

stellung gehe ich davon aus, daß diese explizit dem entspricht, was Haag vage mit dem Ausdruck "Wahl einer entsprechenden Skaleneinheit" umschreibt.

1) Ausführliche Erläuterungen zu den verschiedenen empirischen Skalierungsverfahren finden sich bei Haag, S.87ff.

Würden wir das jeweilige Skalierungsintervall nicht mit dem Tatschuldrahmen in einem konkreten Fall gleichsetzen, so könnten wir mit einer günstig gewählten Skalierungseinheit sogar astronomische Zahlen für die tatschuldangemessene Strafe herleiten.

Sei etwa das Skalierungsintervall für jedes x_i von S.394 [0, 3]. Ferner seien $x_1 = x_2 = x_3 = x_4 = 3$, $a_1 = a_2 = \dots = 0$ und $S_a = 30$, $k = 1$. Wir erhalten somit

$$t = 81 = \log \frac{S_t}{30}.$$

Daraus folgt $S_t = 5 \cdot 10^{33}$ Tage.

Will man die Tatschuld allerdings nicht (wie oben) global bestimmen, sondern sie (was wünschenswert wäre und was Haag eigentlich auch will) aus den einzelnen Tatschuldkomponenten ermitteln, so taucht hier das Problem auf, geeignete Skalierungsintervalle für die einzelnen Tatschuldkomponenten, insbesondere für die strafändernden Merkmale, anzugeben, um ein mit der obigen globalen Bestimmung der Tatschuld aus dem Tatschuldrahmen konsistentes Ergebnis zu erhalten (jene entsprechen offensichtlich (im obigen Beispiel) nicht dem Tatschuldrahmen [0, 4.1]). Man kann sogar vermuten, daß es überhaupt kein Verfahren für die Konstruktion solcher Skalierungsintervalle gibt. Aus diesem Grund scheint Haags Methode - soweit es sich um eine nichtglobale Tatschuldbestimmung handelt - weniger befriedigend zu sein.

Im folgenden wird anstelle einer empirischen Methode für die Bestimmung der Tatschuld ein rechnerisches Verfahren für die Bestimmung der tatschuldangemessenen Strafe angegeben. Bei diesem erfolgt die Berechnung der tatschuldangemessenen Strafe direkt aus der Normalstrafe unter Berücksichtigung der strafbegründenden (diese werden multiplikativ auf die obere Grenze des Strafrahmens angerechnet) und der strafmodifizierenden Merkmale (diese werden additiv auf die Normalstrafe angerechnet).

4.3. Ein neues Strafzumessungsmodell

Im folgenden wird das Modell von Haag modifiziert und anschließend eine rechnerische Methode zur Bestimmung der tatschuldangemessenen Strafe angegeben.

Modifikationen:

(1) Bei der rationalen Strafzumessung (Entscheidungssituation unter Sicherheit) kommen nur Straf- bzw. Geldhöhen als mögliche Maßnahmen in Betracht. Alle anderen von Haag angeführten Maßnahmen s1-s3 (als Strafrahen), n1-n5, m1-m8 gehören in das Entscheidungsmodell unter Risiko.

(2) Wir lassen alle strafrechtlich relevanten Ziele im Sinne der im 2.Kap. dieses Teils entwickelten Theorie des juristischen Nutzens zu.

(3) Bei der Beurteilung der Normal- und tatschuldangemessenen Strafe berücksichtigen wir die Wahrscheinlichkeit, mit der der strafbare Tatbestand festgestellt ist.

(4) Wir berücksichtigen außerdem neben der Schwere der Tat auch die Schwere der Schuld. (Die Schuld muß dabei mit überwiegender Wahrscheinlichkeit nachgewiesen werden; vgl.S.369f dieser Arbeit).

Bei der Berechnung der Strafe wollen wir jedoch nur den Wahrscheinlichkeitswert für den objektiven Sachverhalt berücksichtigen. Ferner wollen wir die Schwere des (sozialen bzw. individuellen) Vorwurfs¹⁾ außer Acht lassen, denn diese wird ohnehin entweder durch den gesetzlichen Strafrahen oder durch die strafmodifizierenden Gründe erfaßt.

Ein 4-Tupel $\langle M, Z, g, w \rangle$ ist genau dann ein rationales Strafzumessungsmodell, wenn gilt:

- (1) M ist eine nicht leere und endliche Menge von konkretisierten Maßnahmen; diese kürzen wir mit $m_1, \dots, m_n$ ab. (M ist somit identisch mit dem gesetzlichen Strafrahen bzw. einer Maßnahme aus dem 2.Kap.; m_1 ist die Höhe einer Geld- oder Gefängnisstrafe).

1) Siehe S.394 dieser Arbeit.

Für alle $m_i \in M$ gilt:

$$m_u \leq m_i \leq m_o \quad \text{und} \quad m_i \leq m_t \quad m_u, m_o, m_t \in M$$

(m_u ist die untere Grenze des Strafrahmens M ; m_o die obere Grenze des Strafrahmens M , m_t die tatschuldangemessene Strafe aus M)¹⁾

- (2) Z ist eine nicht leere und endliche Menge, deren Elemente Funktionen (Zielfunktionen) von M in $[0,1]$ sind. Die Elemente von Z (Ziele) bezeichnen wir mit $Z_1, \dots, Z_r$.
- (3) g ist eine Funktion ("Gewichtsfunktion") von Z in $[0,1]$; g ordnet jedem $Z_j \in Z$ eine reelle Zahl q mit $0 \leq q \leq 1$, sein Gewicht, zu.
- (4) w ist eine Funktion von M in $\mathbb{R}$, für die gilt:

$$\text{für alle } m_i \in M \quad w(m_i) = \sum_{j=1}^r g(Z_j) \cdot Z_j(m_i).$$

Die mit diesem Modell verbundene normative Regel besagt:

Wird das Entscheidungssubjekt durch das Modell $\langle M, Z, g, w \rangle$ charakterisiert, so soll es ein $m \in M$ wählen, für das $w(m) \rightarrow \max$.

Wie wir sehen, sind die Bedingungen (2), (3) und (4) in diesem Modell mit den Bedingungen (2), (4) und (5) im Haag-Modell identisch. Die Bedingungen (1) und (3) des Haag-Modells sind hingegen in eine einzige Bedingung (1) transformiert, wobei unter den Maßnahmen diesmal nicht (wie bei Haag) alle möglichen Maßnahmen verstanden werden, sondern lediglich konkretisierte Maßnahmen (Straf- bzw. Geldhöhe) aus einem bestimmten Strafrahen. Die Menge solcher konkretisierten Maßnahmen wird mit dem Strafrahen (d.h. mit der "Maßnahme" im Sinne des 2.Kap.) identifiziert. Da die Strafen, die der unteren, der oberen und der tatschuldangemessenen Strafe entsprechen und die Wahl einer konkretisierten Maßnahme aus dem Strafrahen einschränken²⁾, auch in diese Menge gehören, erübrigt sich die explizi-

1) Auch hier werden m_i , m_u , m_o und m_t nicht als solche, sondern als auf positive ganze Zahlen abgebildet betrachtet.

2) wobei die Bedingung $m_u \leq m_i \leq m_o$ trivialerweise gilt und deswegen weggelassen werden kann, denn die Menge M (d.h. der gesetzliche Strafrahen) ist gerade durch die Angabe ihrer kleinsten und größten Elemente charakterisiert. Allerdings wird im StGB die untere bzw. die obere Grenze des Strafrahmens nicht immer explizit angeführt. In solchen Fällen regelt §38 StGB bei zeitigen Freiheitsstrafen die Dauer einer solchen Strafe; vgl.: "Das Höchstmaß der zeitigen Freiheitsstrafe ist fünfzehn Jahre, ihr Mindestmaß ein Monat".

te Anführung der Restriktionsmenge $R^1)$

Bestimmung der tatschuldangemessenen Strafe

Wir verwenden folgende Abkürzungen:

m_u, m_o	unter bzw. obere Grenze des Strafrahmens
m_m	mittlere bzw. Normalstrafe
m_a	Strafe aus dem Strafrahmen, bei der die Belastung am geringsten empfunden wird (es ist zu erwarten, daß eine solche gleich der unteren Grenze des Strafrahmens ist)
p	Wahrscheinlichkeitswert des strafbaren Sachverhalts (d.h. des objektiven Tatbestands)
s_t	Schwere der Tat ²⁾³⁾
s_c	Schwere der Schuld ⁴⁾
m'_o	modifizierte obere Grenze
m_t	tatschuldangemessene Strafe

p, s_t, s_c sind auf das Intervall $[0,1]$ normiert.

1) Im Haag-Modell haben wir zwei Mengen S und R unterschieden, d.h. zwischen derjenigen aller möglichen Maßnahmen und derjenigen aller möglichen Restriktionen (wir betrachteten die beiden Mengen nicht als solche, sondern als auf eine endliche Menge positiver ganzer Zahlen abgebildet). Eine solche Unterscheidung war damals unumgänglich, denn ein und dieselbe Maßnahme aus S (dargestellt durch eine bestimmte positive ganze Zahl) konnte verschiedenen Strafrahmens (d.h. verschiedenen unteren und oberen Grenzen) genügen und außerdem verschiedenen Straftaten tatschuldangemessen sein.

2) Die "Schwere der Tat" soll so verstanden werden, daß z.B. ein Diebstahl von 20 DM "objektiv gesehen" als leichter angesehen wird als ein Diebstahl von 1000 DM (Es ist allerdings fraglich, ob eine solche Klassifizierung ohne Bezugnahme auf die jeweilige Person und ihre wirtschaftlichen Verhältnisse ausreicht).

3) Zwar erscheint es bei manchen Delikten, wie Diebstahl, Körperverletzung, usw. sinnvoll, von der Schwere der Tat innerhalb des Strafrahmens zu reden, jedoch nicht bei verschiedenen Tötungsverbrechen (auch wenn das StGB unterschiedliche Strafbemessungsmaßstäbe je nach Tötungsart, Absicht usw. kennt). Denn das Töten eines anderen Menschen (als objektive Handlung) ist und bleibt die höchste Verletzung und wiegt daher - unabhängig davon, ob irgendwelche Rechtfertigungsgründe dafür vorliegen oder nicht - in jedem Fall gleichermaßen schwer. Hier könnte man daher stets $s_t = 1$ ansetzen. Was hingegen die "Schwere der Schuld" bei Tötungsverbrechen betrifft, so ist hier eine Differenzierung möglich. Allerdings wird die Schwere der Schuld gerade bei solchen Verbrechen bereits durch entsprechende Strafrahmens berücksichtigt.

4) Gemeint ist hier der Grad des Verschuldens des Täters, d.h. Vorsatz, Fahrlässigkeit und Mitverschulden des Opfers. Normiert man die

Berechnung der mittleren (oder Normal-) Strafe:

Zuerst wird die obere Grenze des Strafrahmens modifiziert:

$$m'_0 = m_0 \cdot p \cdot s_t \cdot s_c.$$

Sodann die mittlere Strafe (analog wie bei Haag) errechnet:

$$m_m = \sqrt{m_u \cdot m'_0}.$$

(Wir gehen also davon aus, daß der Zusammenhang zwischen einer Strafe und der vom Täter empfundenen Belastung logarithmisch ist und durch das Fechnerische Gesetz dargestellt werden kann (vgl. S. 392 dieser Arbeit), wohlwissend, daß es sich dabei um eine Hypothese handelt¹⁾).

Weitere Bezeichnungen:

s' Abstand der mittleren Strafe von der modifizierten oberen Grenze mit $s' = \sum_{i=1}^n d_i s'$, d.h. $\sum_{i=1}^n d_i = 1$; die d_i sind Gewichte, die man einzelnen strafferhöhenden Zumessungsgründen zuordnet.

s'' Abstand der mittleren Strafe von der unteren Grenze mit $s'' = \sum_{j=1}^l e_j s''$, d.h. $\sum_{j=1}^l e_j = 1$; die e_j sind Gewichte, die man einzelnen strafmildernden Zumessungsgründen zuordnet.

Es gilt:

$$\begin{aligned} s' &= m'_0 - m_m = \sum_{i=1}^n d_i s' \\ s'' &= m_m - m_u = \sum_{j=1}^l e_j s'' \end{aligned}$$

Die Formel für die tatschuldangemessene Strafe:

$$m_t = m_m + \sum_{i=1}^k d_i s' - \sum_{j=1}^r e_j s'', \quad k \leq n, \quad r \leq l.$$

(Die oberen Indizes k und r können jeden Wert von 0 bis n bzw. l

Schwere der Schuld auf das Intervall $[0,1]$, so ist es sinnvoll, s_c wie folgt anzusetzen: s_c (Vorsatz) = 1, s_c (Fahrlässigkeit) $\in [1/2, 1]$, s_c (Mitverschulden beim Vorsatz) $\in [1/2, 1]$ (d.h. die Schwere der Schuld bei einer vorsätzlich begangenen Handlung reduziert sich auf die Hälfte bei gravierendem Mitverschulden des Opfers; bei Fehlen des Mitverschuldens des Opfers ist $s_c = 1$; ansonsten liegt s_c bei Mitverschulden zwischen 1/2 und 1); s_c (Mitverschulden bei Fahrlässigkeit) $\in [0, 1]$ (d.h. die Schwere der Schuld bei einer fahrlässig begangenen Handlung kann sich bei Mitverschulden des Opfers evtl. bis auf 0 reduzieren; bei Fehlen eines solchen kann sie evtl. auch die Grenze 1 erreichen). Mitverschulden des Opfers ohne Vorsatz oder Fahrlässigkeit seitens des Täters wird nicht beachtet, denn in diesem Fall ist offensichtlich $P(\text{Schuld}) \leq 1/2$.

1) Auch wenn eine solche Hypothese plausibel erscheint, bedarf sie trotzdem einer eingehenden Überprüfung; vgl. Haag, S. 62ff.

annehmen; ist z.B. $k = 0$ (d.h. es liegt kein strafferhöhendes Merkmal vor) und $r = 1$ (d.h. es liegen alle möglichen strafmildernden Merkmale vor), so ist $m_t = m_m - s'' = m_m - m_m + m_u = m_u$ (d.h. gleich der unteren Grenze oder der geringsten Strafe aus dem Strafrahmen); ist dagegen $k = n$ (d.h. es liegen alle möglichen strafferhöhenden Merkmale vor) und $r = 0$ (d.h. es liegt kein einziges strafmilderndes Merkmal vor), dann ist $m_t = m_m + s' = m_m + m'_0 - m_m = m'_0$ (d.h. m_t erschöpft völlig die modifizierte obere Grenze des Strafrahmens); ist $0 < k < n$, $0 < r < 1$, so bedeutet dies, daß nur einige strafferhöhende und strafmildernde Merkmale vorliegen¹⁾

Zur Veranschaulichung

Wir nehmen folgende Werte an:

$m_u = 30$ Tage

$m_o = 4$ Jahre bzw. 1461 Tage

$p = 0.8$ (= 0.6)

$s_c = 1$ (d.h. die Tat ist vorsätzlich begangen; vgl. S. 399f, Fn. 4).

Je nach Schwere der Tat kann man folgende Strafen errechnen²⁾

s_t	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
$p = 0.8, m'_0$	117T	234T	351T	468T	584T	701T	818T	935T	1052T	1168T
m_m	59T	84T	103T	118T	132T	145T	157T	168T	178T	187T
$p = 0.6, m'_0$	88T	175T	273T	351T	438T	526T	614T	701T	789T	877T
m_m	51T	73T	89T	103T	115T	126T	136T	145T	154T	162T

Tabelle 1

Wir wollen nun eine tatschuldangemessene Strafe für $p = 0.8$ und $s_t = 0.9$ (d.h. $m_m = 178$ Tage in der obigen Tabelle) errechnen.

Angenommen, es gibt insgesamt 5 strafferhöhende und 3 strafmildernde

1) $s' = \frac{n}{i} d_i s'$ besagt z.B., daß bei Vorliegen aller strafverschärfenden Merkmale die Strafe bis zur modifizierten oberen Grenze ausgeschöpft werden muß; $s' = \frac{k}{i} d_i s'$ mit $k < n$ besagt dagegen, daß nur k von n strafverschärfenden Merkmalen vorliegen; die so erhaltene Strafe ist natürlich kleiner als diejenige, bei der alle strafferhöhenden Merkmale vorliegen (oben gibt es also insgesamt n strafferhöhende und l strafmildernde Merkmale).

2) Die Zahlen werden im folgenden immer abgerundet.

Gründe (als erschöpfende Merkmale)¹⁾ Zur Vereinfachung nehmen wir weiter an, daß alle Strafzumessungsgründe gleiches Gewicht haben, d.h. daß jedes strafe erhöhende Merkmal in diesem Fall die mittlere Strafe um

$$\frac{m'_0 - m_m}{5} = \frac{1052 - 178}{5} = 175 \text{ Tage}$$

erhöht und jedes strafmildernde Merkmal die mittlere Strafe um

$$\frac{m_m - m_u}{3} = \frac{178 - 30}{3} = 49 \text{ Tage}$$

verringert²⁾

Es gibt dann 23 Möglichkeiten, eine Strafe zu bilden (vgl. die nachfolgende Tabelle).

1) Natürlich kann die Anzahl der strafändernden Merkmale anders sein. Es ist Aufgabe des Gerichts, eine Liste aller möglichen strafmodifizierenden Merkmale und ihrer Gewichtung auszuarbeiten. Diese Liste muß außerdem für jeden einer bestimmten Straftat angeklagten immer die gleiche sein. Sie variiert lediglich hinsichtlich der Straftat selbst (wie etwa "Mord", "Totschlag", "Diebstahl" usw.). Die obige Annahme ist somit so zu verstehen, daß eine Liste mit insgesamt 8 strafmodifizierenden Merkmalen für alle Fälle einer bestimmten Straftat festgelegt ist und in einem konkreten Fall überprüft wird. Zu tatschuldmodifizierenden Gründen vgl. §46 (2) StGB (mit Ausnahme des Vorlebens des Täters, das mit der Tatschuld nichts zu tun hat). Zu beachten ist, daß die Berücksichtigung der Strafzumessungsgründe bei der Bestimmung der Tatschuld nicht nach präventiven Gesichtspunkten erfolgen darf. Die Tatschuld hängt einzig und allein von der Tat und dem Verschulden des Täters ab, nicht dagegen von sonstigen, mit der Tat nicht unmittelbar zusammenhängenden Merkmalen und Erwägungen. Allerdings können solche Gesichtspunkte unterhalb der tatschuldangemessenen Grenze berücksichtigt werden; dazu später auf S.405 dieser Arbeit.

2) In der Regel haben strafmodifizierende Merkmale nicht alle das gleiche Gewicht, so daß das Vorliegen verschiedener Merkmale die Strafe nicht um die gleiche Anzahl von Tagen erhöht bzw. verringert. Diesem Umstand trägt die obige Methode Rechnung. Zur Veranschaulichung: Sei bei den oben angenommenen 3 strafmildernden Merkmalen die Gewichtsverteilung (1/6, 2/3, 1/6) gegeben, d.h. das zweite Merkmal ist doppelt so wichtig wie die beiden anderen zusammengekommen. In diesem Fall verringert das Vorliegen des ersten bzw. des dritten Merkmals die Strafe jeweils um

$$\frac{178 - 30}{6} = 25 \text{ Tage,}$$

das Vorliegen des zweiten Merkmals um

$$\frac{2(178 - 30)}{3} = 99 \text{ Tage.}$$

	straferhöhende					strafmildernde			
	Merkmale								
	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{3}{5}$	$\frac{4}{5}$	$\frac{5}{5}$	$\frac{1}{3}$	$\frac{2}{3}$	$\frac{3}{3}$	m_t
1.	x								353 T
2.		x							528 T
3.			x						703 T
4.				x					878 T
5.					x				1052T
6.						x			129 T
7.							x		80 T
8.								x	30 T
9.	x					x			304 T
10.	x						x		255 T
11.	x							x	206 T
12.		x				x			479 T
13.		x					x		430 T
14.		x						x	381 T
15.			x			x			654 T
16.			x				x		605 T
17.			x					x	556 T
18.				x		x			829 T
19.				x			x		780 T
20.				x				x	731 T
21.					x	x			1003T
22.					x		x		954 T
23.					x			x	905 T

Tabelle 2

(In der obigen Tabelle besagt z.B. die Zeile 16, daß drei strafferhöhende und zwei strafmildernde Merkmale vorliegen; die tatschuldangemessene Strafe m_t beträgt somit 605 Tage; die Zeile 5 drückt aus, daß alle (insgesamt fünf) strafferhöhenden Merkmale und kein einziges strafmilderndes Merkmal vorliegen; die tatschuldangemessene Strafe ist hier gleich der modifizierten oberen Grenze (1052 Tage; vgl. Tab. 1); die Zeile 8 dagegen sagt aus, daß alle strafmildernden und kein einziges strafferhöhendes Merkmal vorliegen; die tatschuldangemessene Strafe ist hier somit gleich der unteren Grenze des Strafrahmens)

Die hier vorgeschlagene Methode, bei der zunächst die obere Grenze des vorgegebenen Strafraumens durch die Berücksichtigung der Schwere der Tat, der Schwere der Schuld und der Wahrscheinlichkeit für den objektiven Tatbestand modifiziert wird und erst dann die Berechnung der tatschuldangemessenen Strafe erfolgt, lehnt sich an den Vorschlag von Dubs¹⁾ an, bei der Strafzumessung zuerst eine Normalstrafe nach der "Schwere der Tat" und erst dann durch die Berücksichtigung aller strafmodifizierenden Faktoren die individuelle tatschuldangemessene Strafe zu bestimmen.

Im Unterschied zu Dubs, der bei der Bestimmung der Normalstrafe nur die "Schwere der Tat" berücksichtigt, haben wir nicht nur diese, sondern auch die Wahrscheinlichkeit für den objektiven Tatbestand wie auch die "Schwere der Schuld" auf die obere Grenze des Strafraumens angerechnet. Die Berücksichtigung der Wahrscheinlichkeit für den objektiven Tatbestand bei der Konkretisierung der Strafe dient der Verhinderung dessen, daß unabhängig davon, mit welcher Wahrscheinlichkeit der strafbare Tatbestand festgestellt ist (diese kann 0.51 oder 0.99 betragen), ein und dieselbe konkretisierte Maßnahme gewählt wird. (Nach unserer Methode ergibt eine geringere Wahrscheinlichkeit für den Tatbestand eine entsprechend geringere konkrete Strafhöhe). Die Berücksichtigung der Schwere der Schuld soll darüberhinaus verhindern, daß bei einem leichteren Verschulden (wenn z.B. Fahrlässigkeit oder ein mehr oder weniger gravierendes Mitverschulden des Opfers vorliegen) die gleiche konkretisierte Maßnahme gewählt wird wie bei Vorsatz oder schwerem Verschulden.

Grassberger²⁾ dagegen geht von der Normalstrafe aus, auf die dann alle strafmodifizierenden Merkmale, die die Tatschuld beeinflussen, darunter auch die Schwere der Tat und die der Schuld, angerechnet werden. Würden wir nach Grassbergers Vorschlag verfahren, so bestünde unsere Methode darin, zuerst im Strafraumen die "Normalstrafe" zu bestimmen und dann auf diese multiplikativ die Schwere der Tat, die Schwere der Schuld und die Wahrscheinlichkeit für den objektiven Tatbestand anzurechnen. Auf die so erhaltenen durchschnittlichen tatschuldangemessenen Strafen würden wir schließlich (wie im Fall der ersten Methode) alle individualisierenden strafmodifizierenden Merkmale additiv anrechnen³⁾.

Welche Methode (diese oder die im Text vorgeschlagene) sich für die praktische Strafzumessung geeigneter erweist, ist eine empirische Frage und kann hier nicht beantwortet werden.

Gegen die zweite Methode spricht allerdings, daß sie einerseits bei der Berechnung der durchschnittlichen tatschuldangemessenen Strafe (Normalstrafe) "zu milde" ist. (Die mit dieser Methode berechneten mittleren Strafen liegen unter den mit der ersten Methode bestimmten Normalstrafen; so erhalten wir in unserem Beispiel keine Strafe, wenn die Straftat mit einer Wahrscheinlichkeit von 0.8 festgestellt ist, die Schwere der Tat aber 0.1 beträgt. Dann beträgt die mittlere tatschuldangemessene Strafe nur 17 Tage, was unterhalb der unteren Grenze des Strafraumens von 30 Tagen liegt. (Mit der ersten Methode erhalten wir hier hingegen eine Strafe von 59 Tagen). Ist die Wahrscheinlichkeit für den Tatbestand 0.6, so erhalten wir keine Strafe auch dann, wenn die Schwere der Schuld bei 0.2 liegt. (Mit der ersten Methode beträgt die Strafdauer bei $p = 0.6$ und $s_t = 0.2$ 73 Tage)). Andererseits ist sie "zu hart", wenn es um die Bestimmung der individuellen tatschuldangemes-

1) H. Dubs, Analytische Bewertung als Grundlage richterlicher Strafzumessung. Festgabe zum Schweizerischen Juristentag (1963), S.22.

2) R. Grassberger, Die Strafzumessung (1932); S.80ff.

3) Die Formel lautet:

$$m_t = m'_m + \sum_{i=1}^k d_i s^i - \sum_{j=1}^r e_j s^j, \quad k \leq n, \quad r \leq l$$

mit $m'_m = \sqrt{m_u \cdot m_o \cdot p \cdot s \cdot s_c}$ und s^i als Abstand von m'_m von der oberen Grenze des Strafraumens und s^o dem von der unteren. (m'_m ist die mittlere tatschuldangemessene Strafe; sonstige Bezeichnungen wie im Text).

senen Strafe geht, denn nach ihr kann auch die obere Grenze des Strafrahmens allein durch die Berücksichtigung von strafmodifizierenden Merkmalen selbst bei geringen Wahrscheinlichkeitswerten für den Tatbestand bei Vorliegen leichter Straftaten und bei einem leichten Verschulden voll ausgeschöpft werden.

Berücksichtigung von general-, spezialpräventiven und sonstigen Erwägungen

Zunächst sei noch einmal angemerkt, daß die Bestrafung des Täters die tatschuldangemessene Maßnahme weder aus general- bzw. spezialpräventiven noch aus sonstigen Erwägungen übersteigen darf. Außerdem darf die Bestimmung der eigentlichen Tatschuld, welche sich aus dem Unrecht und dem Verschulden einer konkreten Straftat ergibt, nicht nach präventiven Gesichtspunkten (die mit der Straftat nichts zu tun haben) erfolgen.

Die Berücksichtigung präventiver Gesichtspunkte und Erwägungen kann also nur zu einer mildernden als der tatschuldangemessenen Bestrafungspraxis führen, nicht dagegen zu einer härteren. Wir können diesen Gedankengang etwas konkreter ausdrücken: Die in §48 StGB (Rückfall) und teilweise in §46 StGB¹⁾ (Grundsätze der Strafzumessung) angeführten strafmodifizierenden Merkmale dürfen nur positiv (strafmildernd)²⁾ bzw. höchstens tatschuldausschöpfend³⁾, jedoch nicht negativ (d.h. tatschuldkonstruierend oder über die Tatschuld hinaus strafferhöhend)⁴⁾ berücksichtigt werden. In der Entscheidungssituation unter Sicherheit wird somit stets die tatschuldangemessene mit mildernden Maßnahmen unter dem Gesichtspunkt der maximalen Zielerreichung verglichen⁵⁾.

1) in bezug auf das Vorleben des Täters, welches bei der Bestimmung der Tatschuld nicht berücksichtigt wird.

2) Etwa, wenn vom Täter kein Rückfall zu erwarten ist bzw. wenn sein Vorleben voller negativer Erfahrungen war, etc.

3) falls z.B. nach resozialisierenden Gesichtspunkten die Strafe in einem konkreten Fall viel milder als tatschuldangemessen ausfallen würde, es aber zweckmäßig ist, die Strafe aus generalpräventiven oder sonstigen Erwägungen doch wiederum bis zur tatschuldangemessenen zu erhöhen.

4) Die Rückfallerwartung oder eine besonders verwerfliche Lebensart dürfen sich somit nicht über die Tatschuld hinaus als strafferhöhend auswirken; sie dürfen höchstens innerhalb des von der Tatschuld begrenzten Strafrahmens berücksichtigt werden; vgl. Haag, S.21: "Die Rechtssprechung erkennt die Generalprävention als Strafzumessungsgrund sowie als Strafzweck in den §23, 276 StGB grundsätzlich an".

5) Beachte: Alle sonstigen strafferhöhenden und -mildernden Merkmale (außer präventiven und resozialisierenden) werden bereits bei der Bestimmung der tatschuldangemessenen Maßnahme berücksichtigt.

Beispiel:

Es liege ein (aus spezialpräventiven Erwägungen möglicher) Straf-
milderungsgrund vor. Nach §49 StGB darf bei einer zeitlich begrenz-
ten Freiheitsstrafe höchstens auf $3/4$ des angedrohten Höchstmaßes
erkannt werden, falls besondere Milderungsgründe vorliegen. Die Ent-
scheidung soll unter Berücksichtigung der folgenden drei Zielsetzun-
gen erfolgen: Z_1 : angemessener Ausgleich des Übels (Gerechtigkeit),
 Z_2 : Verbrechensverhütung mittels Resozialisation und Z_3 : Vermeidung
von Drittschäden. Ihre relativen Bedeutungen (Gewichte) seien jeweil
 $g(Z_1) = 0.4$, $g(Z_2) = 0.4$, $g(Z_3) = 0.2$; d.h. Z_1 und Z_2 sind in die-
sem Fall gleich wichtig und doppelt so wichtig wie Z_3 .
Die folgende Tabelle möge diese Entscheidungssituation veranschau-
lichen:

	0.4 Z_1	0.4 Z_2	0.2 Z_3
m_t	1	0.3	0
$m_{<t}$	0	0.7	1

$$w(m_t) = 1 \cdot 0.4 + 0.3 \cdot 0.4 + 0 \cdot 0.2 = 0.52$$

$$w(m_{<t}) = 0 \cdot 0.4 + 0.7 \cdot 0.4 + 1 \cdot 0.2 = 0.48.$$

Somit ist m_t die optimale Maßnahme (zu ihrer Berechnung vgl. S.399ff
dieser Arbeit).

(in der obigen Tabelle bedeutet: m_t die tatschuldangemessene Maßnahme; $m_{<t}$ eine
mildere als die tatschuldangemessene Maßnahme, $Z_i(m_t)$ das Maß der Wirksamkeit
von m_t zur Erreichung des Ziels Z_i , $i = 1, 2, 3$; $Z_i(m_{<t})$ das Maß der Wirksamkeit
von $m_{<t}$ zur Erreichung des Ziels Z_i , $i = 1, 2, 3$; $w(m_t)$ die Gesamtwirksamkeit von
 m_t zur Erreichung aller Ziele; $w(m_{<t})$ die Gesamtwirksamkeit von $m_{<t}$ zur Erreichung
aller Ziele).

(Falls $m_{<t}$ die optimale Maßnahme wäre, so hätte man die tatschuld-
angemessene Grenze des Strafrahmens höchstens mit $3/4$ ¹⁾ multipli-
zieren und in dem einerseits von dieser neuen 'oberen' Grenze, an-
dererseits von der nach §49 (3) StGB²⁾ modifizierten unteren Grenze
beschränkten Strafrahmen die optimale Maßnahme wählen müssen³⁾).

1) §49 StGB (Besondere gesetzliche Milderungsgründe) besagt u.a.:
"... 2. Bei zeitiger Freiheitsstrafe darf höchstens auf drei Viertel des angedroh-
ten Höchstmaßes erkannt werden. Bei Geldstrafe gilt dasselbe für die Höchstzahl
der Tagessätze".

2) §49 (3) StGB lautet: "Das erhöhte Mindestmaß einer Freiheitsstrafe ermä-
ßigt sich im Falle eines Mindestmaßes von zehn oder fünf Jahren auf zwei Jahre,
im Falle eines Mindestmaßes von drei oder zwei Jahren auf sechs Monate, im Falle
eines Mindestmaßes von einem Jahr auf drei Monate, im übrigen auf das gesetzli-
che Mindestmaß".

3) etwa durch die von Haag vorgeschlagenen Methoden der empirischen
Ermittlung des funktionalen Zusammenhangs zwischen einer Strafe und
ihrer Wirksamkeit zur Erreichung eines bestimmten Ziels; vgl.
Haag, S.185ff.

Anhang

I. Metrisierungsmethoden

Ein Jurist wird sich zu Recht fragen, wie er zu seinen Nützlichkeits- bzw. Wahrscheinlichkeitswerten kommt. Die Theorie der Metrisierung bietet ihm als Antwort darauf verschiedene Methoden, die sowohl in ihrem Exaktheitsgrad als auch im Plausibilitätsgehalt unterschiedlich sind. Im folgenden wird keine Metrisierungstheorie entwickelt (eine solche nimmt v.a. bei Fishburn einen breiten Raum ein¹⁾), sondern es wird lediglich kurz angegeben, worin diese Metrisierungsmethoden bestehen²⁾.

I.1. Metrisierung von Nützlichkeiten

Man unterscheidet gewöhnlich zwischen folgenden Metrisierungsmethoden (die hier nach steigendem Exaktheitsgrad und fallendem Plausibilitätsgehalt aufgeführt sind):

partielle Ordinalmetrisierung

ordinale Metrisierung

Metrisierung erster Ordnung

Metrisierung n-ter Ordnung, $n = 2, 3, \dots$

Metrisierung höherer Ordnung

beschränkte Intervallmetrisierung erster Stufe

beschränkte Intervallmetrisierung zweiter Stufe

beschränkte Intervallmetrisierung dritter Stufe

Intervallmetrisierung

Im folgenden benutzen wir die Bezeichnungen:

$\mathcal{R}$ ist die Menge der Konsequenzen

$R_1, \dots, R_r$ sind Elemente von $\mathcal{R}$, d.h. Konsequenzen³⁾

$u_1, \dots, u_r$ die ihnen entsprechenden relativen Werte (Nützlichkeiten bzw. Wünschbarkeiten); diese werden durch reelle Zahlen repräsentiert

" $R_i \succcurlyeq R_j$ ": " R_j wird R_i nicht vorgezogen" bzw. " R_i ist mindestens so wünschenswert wie R_j "

1) P.C.Fishburn, Decision and Value Theory (1964), v.a. 4. und 5. Kap., an die sich die folgenden Ausführungen anlehnen.

2) Der Leser soll sich nicht von der folgenden formalen Ausführungsweise abschrecken lassen, denn es handelt sich hier im Grunde genommen um triviale Sachverhalte, die aber gerade für einen Anwender der rationalen juristischen Entscheidungstheorie von großem Interesse sein dürften.

3) Im Unterschied zum Haupttext, wo die Konsequenzen durch zwei untere Indizes charakterisiert wurden (welche die Abhängigkeit einer Konsequenz von einem Sachverhalt und einer Maßnahme symbolisieren sollten), werden die Konsequenzen hier nur durch einen einzigen unteren Index, der die Stelle einer Konsequenz in der Präferenzordnung der Konsequenzen angibt, gekennzeichnet. Insgesamt gibt es also r Konsequenzen. Auch die Bezeichnung " u_i , $i=1, \dots, r$ " ist von der Bezeichnung " u_{ij} , $i=1, \dots, m; j=1, \dots, n$ " zu unterscheiden; der Index i bei " u_i " steht für die Stelle

" $R_i \sim R_j$ ": " R_i und R_j sind gleichwertig" bzw. "das Individuum ist in bezug auf R_i und R_j indifferent"

$\succ$ ist eine Präferenzrelation

$\sim$ ist eine Indifferenzrelation bzw. Äquivalenzrelation

$\geq$ ist eine Relation zwischen reellen Zahlen (zu lesen als: "nicht kleiner als")

Dementsprechend interpretiert man die Relationen: $\leq, <, \leq, >, <, =$.

(1) Ordinale Metrisierung

Diese Methode besteht darin, daß man die relativen Werte (d.h. Nützlichkeiten bzw. Wünschbarkeiten) der Konsequenzen $R_1, \dots, R_r$ nach ihrer Größe ordnet. Es entsteht somit eine Ordnung (Nutzenordnung):

$$u_1 \geq u_2 \geq \dots \geq u_r$$

Für die ordinale Metrisierung gilt das folgende Axiom:

A1 Die Relation $\succ$ ist eine schwache Ordnung

d.h. es gilt:

A1(1) $R_i \succ R_j$ oder $R_j \succ R_i$ für alle $R_i, R_j \in \mathcal{R}$ (d.h. $\succ$ ist strikt konnex)

A1(2) Wenn $R_i \succ R_j$ und $R_j \succ R_k$, dann $R_i \succ R_k$ (d.h. $\succ$ ist transitiv).

Die Beziehung zwischen den Konsequenzen und ihren relativen Werten ist durch die Grundbehauptung:

$$u(R_j) \geq u(R_k) \leftrightarrow R_j \succ R_k$$

charakterisiert.

Die obige Behauptung besagt einfach: wenn eine Konsequenz mindestens so wünschenswert ist wie eine andere, dann ist der relative Wert der ersten mindestens so groß wie der der zweiten, und wenn das nicht der Fall ist, dann ist der relative Wert der ersten Konsequenz kleiner als der der zweiten. (Umkehrung analog).

Die ordinale Metrisierung (oder schwache Ordnung) zwischen Nützlichkeiten

$u_1 \geq u_2 \geq \dots \geq u_r$ ergibt sich also unmittelbar aus der schwachen Präferenzordnung zwischen den Konsequenzen: $R_1 \succ R_2 \succ \dots \succ R_r$, die durch A1 garantiert wird.

(Die ordinale Metrisierung ordnet somit die relativen Werte der Konsequenzen entsprechend den Präferenzen zwischen Konsequenzen; es ist klar, daß mit der Ordinalmetrisierung noch keine festen Werte für $u_1, u_2, \dots, u_r$ verbunden sind).

Empirische Messung: Durch Befragung des Individuums.

Man präsentiert dem Individuum z.B. paarweise die Konsequenzen und läßt es jedes Paar nach der Präferenz ordnen, bis eine vollständige Ordnung anhand geordneter Paare abgeleitet ist, oder man läßt z.B. das Individuum den Konsequenzen Gewichte zuordnen, so daß das größte Gewicht der wünschenswertesten Konsequenz zukommt, usw.; die Konsequenzen werden dann entsprechend ihren Gewichten geordnet.

in der Nutzenordnung, der Index ij bei " u_{ij} " weist hingegen auf die Stelle in der Nützlichkeitsmatrix hin.

(2) Partielle Ordinalmetrisierung

Der Sinn dieser Methode liegt darin, eine Ordnung zwischen Nützlichkeiten auch dann abzuleiten, wenn die Präferenzordnung zwischen den Konsequenzen teilweise unbekannt ist.

Es gilt das folgende Axiom:

A2 Die Relation $\succsim$ ist eine Quasi-Ordnung auf der Menge $\mathcal{R}$.

Eine Quasi-Ordnung ist eine Relation, die reflexiv ($R_i \succsim R_i$) und transitiv, jedoch nicht konnex ist (d.h. es gilt nicht $R_i \succsim R_j$ oder $R_j \succsim R_i$ für alle $R_i, R_j \in \mathcal{R}$).

Beispiel:

Sei eine Quasi-Ordnung von vier Konsequenzen gegeben:

$$R_1 \succsim R_3 \succsim R_4 \quad \text{und} \quad R_2 \succsim R_3 \succsim R_4$$

Es ist kein Vergleich zwischen R_1 und R_2 möglich. Nichtsdestoweniger kann man mit einer zusätzlichen Metrisierung bestimmen, ob $u_1 \geq u_2$ oder $u_2 \geq u_1$ gilt:

Aus den beiden Präferenzordnungen $R_1 \succsim R_3 \succsim R_4$ und $R_2 \succsim R_3 \succsim R_4$ sind zwei Ordnungen zwischen Nützlichkeiten ableitbar: $u_1 \geq u_3 \geq u_4$ und $u_2 \geq u_3 \geq u_4$. Man vergleicht nun die Differenzen zwischen den relativen Werten in beiden Nutzenordnungen. Es sind verschiedene Ergebnisse möglich, wie etwa

$$u_1 - u_3 \geq u_3 - u_4$$

$$u_3 - u_4 \geq u_2 - u_3$$

Die beiden letzten implizieren aber

$$u_1 - u_3 \geq u_2 - u_3 \quad \text{oder} \quad u_1 \geq u_2.$$

D.h. auch dann, wenn das Individuum nicht imstande ist, eine klare Präferenz zwischen R_1 und R_2 auszudrücken, könnte es trotzdem möglich sein, auf $u_1 \geq u_2$ (was $R_1 \succsim R_2$ impliziert) oder auf $u_2 \geq u_1$ (was $R_2 \succsim R_1$ impliziert) mit einer zusätzlichen Metrisierung zu schließen. (Allerdings ist eine solche Rekonstruktion der Nutzenordnung aus der partiellen Ordnung, insbesondere wenn es viele unvergleichbare Konsequenzen gibt, oft unmöglich).

Einleitung zu den beschränkten Intervallmetrisierungen

Es gelten folgende Axiome:

A3(1) Es gibt eine Konsequenz, etwa R_1 , so daß $R_1 \succsim R_j$ für alle j , und es gibt eine Konsequenz, etwa R_r , so daß $R_j \succsim R_r$ für alle j .

A3(2) Es gilt nicht, daß $R_r \succsim R_1$.

Alle beschränkten Intervallmetrisierungen basieren auf dem gleichen Prinzip:

Man ermittelt das Verhältnis zwischen Wertdifferenzen, d.h.

$$Q = \frac{u_a - u_b}{u_c - u_d} \quad u_c - u_d \neq 0,$$

wobei man die obere und die untere Schranke des Intervalls angibt, in dem dieses Verhältnis liegt. Es gilt somit $\alpha \leq Q \leq \beta$. Die beschränkten Metrisierungen werden dann angewandt, wenn es nicht möglich ist, Q genau zu bestimmen.

(3) Beschränkte Intervallmetrisierung erster Stufe

Es gelten die Annahmen A3(1) und A3(2), d.h. $R_1 \succcurlyeq R_j \succcurlyeq R_r$ für alle j , $j = 1, \dots, r$, und $R_1 \succ R_r$.

Diese Methode besteht im Bestimmen des Verhältnisses zwischen den Wertintervallen $(u_j - u_r)$ und $(u_1 - u_r)$ für jedes j . Sei I_j das j -te Verhältnis

$$I_j = \frac{u_j - u_r}{u_1 - u_r}$$

Es folgt: $I_1 = 1$ und $I_r = 0$ und $0 \leq I_j \leq 1$ für $j = 2, \dots, r-1$.

Die Schranken von I_j bezeichnen wir mit α_j und β_j , $0 \leq \alpha_j \leq \beta_j \leq 1$.

Es gilt: $\alpha_j \leq I_j \leq \beta_j$, $j = 2, \dots, r-1$.

Bestimmung von α_j und β_j , $j = 2, \dots, r-1$.

1. Methode (intuitiver Ansatz)

Beispiel:

Drei Konsequenzen R_1, R_2 und R_3 mit $R_1 \succcurlyeq R_2 \succcurlyeq R_3$ seien gegeben. Wir ordnen R_1 und R_3 beliebige Nützlichkeitswerte zu, wie etwa $u(R_1) = 1$ und $u(R_3) = 0$. Wir wollen nun den Abstand $u(R_2) - u(R_3)$ bestimmen. Zur Vereinfachung benutzen wir eine lineare Skala, die in 10 gleiche Teile unterteilt ist

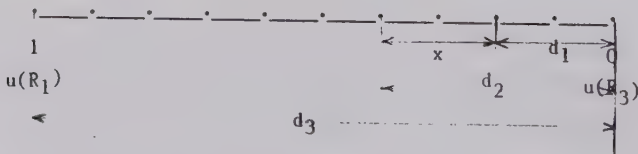


Fig.1.

Das Individuum wird aufgefordert, ein Intervall anzugeben, in dem seinem Gefühl nach der Wert $u(R_2)$ mit Sicherheit enthalten ist, wobei das Intervall möglichst klein sein soll. Angenommen, das Individuum gibt das Intervall x an. Diese Antwort wird folgendermaßen interpretiert: "Das Individuum hält für sicher, daß

$$0.2 \leq \frac{u(R_2) - u(R_3)}{u(R_1) - u(R_3)} \leq 0.4."$$

" $0.2 \leq \frac{u(R_2) - u(R_3)}{u(R_1) - u(R_3)} \leq 0.4$ " ist dabei zu lesen: "Der Abstand $u(R_2) - u(R_3)$ hat einen Wert zwischen 0.2 und 0.4 bzw. der Abstand $u(R_1) - u(R_2)$ hat einen Wert zwischen 0.6 und 0.8" (es ist $u(R_1) - u(R_3) = 1$ nach der obigen Normierung).

Wie dieses Beispiel veranschaulicht, besteht die intuitive Methode in folgendem: Für jedes R_j muß das Individuum auf einer linearen Werteskala mit den Endpunkten $u(R_1)$ und $u(R_r)$ ein Intervall angeben, das seiner Meinung nach den relativen Wert von R_j relativ zu den Werten $u(R_1)$ und $u(R_r)$ enthält. Angenommen, für ein bestimmtes R_j ist das Intervall angegeben, wie Fig.1 es zeigt, mit d_1, d_2 und d_3 als linearen Abständen.

Dann gilt

$$\frac{d_1}{d_3} \leq \frac{u(R_j) - u(R_r)}{u(R_1) - u(R_r)} \leq \frac{d_2}{d_3}.$$

Außer A3(1) und A3(2) gilt für diese Metrisierung eine zusätzliche Annahme:

A4 Für alle R_j , $j = 2, \dots, r-1$, kann das Individuum ein Intervall auf einer linearen Skala mit den Endpunkten $u(R_1)$ und $u(R_r)$ angeben, so daß dieses Intervall $u(R_j)$ enthält.

Mit diesem Intervall für R_j ist

$$\alpha_j \leq I_j \leq \beta_j,$$

mit

$$I_j = \frac{u_j - u_r}{u_1 - u_r}, \quad \alpha_j = \frac{d_1}{d_3}, \quad \beta_j = \frac{d_2}{d_3}$$

wobei d_1, d_2, d_3 die linearen Abstände sind (vgl. Fig.1).

Ziel dieser Methode ist es, eine möglichst kleine Differenz $\epsilon_j = \beta_j - \alpha_j$ bzw. ein möglichst kleines Intervall $[\alpha_j, \beta_j]$ zu erhalten, die/das nach Meinung des Individuums noch immer $u(R_j)$ enthält.

2. Methode (entscheidungs-theoretischer Ansatz)

Der Grundbegriff ist hier der der hypothetischen Lotterien, wie sie in der Nutzentheorie von v. Neumann-Morgenstern auftreten.

Bezeichnungen: : Neben der Menge der Konsequenzen $\mathcal{R}$ die Menge der Lotterien (R_1, p, R_r) für $0 \leq p \leq 1$ und eine Relation $\succsim$ zwischen Elementen dieser beiden Mengen.

Unter einer Lotterie (R_1, p, R_r) versteht man dabei die folgende Wahrscheinlichkeitsverteilung: Man erhält R_1 mit der Wahrscheinlichkeit p und R_r mit der Wahrscheinlichkeit $1-p$.

Das Verfahren basiert auf dem Vergleich zwischen R_j und (R_1, p, R_r) . Wie bei ersten Methode sind es drei Konsequenzen R_1, R_j, R_r , zwischen denen verglichen wird. Die Behauptung $R_j \succsim (R_1, p, R_r)$ ist dabei zu lesen als: "Die Konsequenz R_j wird der Lotterie (R_1, p, R_r) vorgezogen oder ist ihr gleichwertig". Die Interpretation von $(R_1, p, R_r) \succsim R_j$ ist analog.

Außer A3(1) und A3(2) gelten folgende Axiome:

A5(1) Für alle R_j , $j = 2, \dots, r-1$ gibt es einen Wert von p , etwa α_j , mit $R_j \succ (R_1, p, R_r)$ für alle $p \leq \alpha_j$.

A5(2) Es gibt einen Wert von p , etwa β_j , mit $\beta_j \geq \alpha_j$ und $(R_1, p, R_r) \succ R_j$ für alle $p \geq \beta_j$.

A5(3) Wenn $R_j \succ (R_1, p, R_r)$, dann $u_j \geq pu_1 + (1-p)u_r$; und wenn $(R_1, p, R_r) \succ R_r$, dann $pu_1 + (1-p)u_r \geq u_j$.

Aus A3 und A5 folgt

$$\frac{u_j - u_r}{u_1 - u_r} \geq p \quad \text{für alle } p \leq \alpha_j$$

oder

$$p \leq I_j \quad \text{für alle } p \leq \alpha_j,$$

insbesondere

$$\alpha_j \leq I_j.$$

Analog erhält man

$$I_j \leq p \quad \text{für alle } p \geq \beta_j,$$

insbesondere

$$I_j \leq \beta_j.$$

Daraus ergibt sich

$$\alpha_j \leq I_j \leq \beta_j.$$

Das Individuum muß somit zwei Werte von p (d.h. α_j und β_j) angeben mit $R_j \succ (R_1, p, R_r)$ für alle $p \leq \alpha_j$ und $R_j \prec (R_1, p, R_r)$ für alle $p \geq \beta_j$ und $\alpha_j \leq \beta_j$.

Auch hier ist es wünschenswert, daß das Intervall $\epsilon_j = \beta_j - \alpha_j$ möglichst klein ist.

Beispiel:

Die folgende Konsequenzenmatrix

$$\begin{array}{cc} & \begin{array}{c} U_p \\ \bar{U}_p \end{array} \\ \begin{array}{c} H_1(p) \\ H_2 \end{array} & \begin{bmatrix} R_1 & R_3 \\ R_j & R_j \end{bmatrix} \end{array} \quad \begin{array}{l} U_p, \bar{U}_p \text{ sind Umstände, } H_1 \text{ hypothetische Handlungen,} \\ i = 1, 2; P(U_p) = P(R_1) = P(R_j) = p; P(\bar{U}_p) = P(R_j) \\ = 1-p \end{array}$$

charakterisiert eine hypothetische Wahlsituation¹⁾

Die hypothetischen Erwartungswerte der beiden Handlungen lauten:

$$E_u(H_1) = pu_1 + (1-p)u_3,$$

$$E_u(H_2) = pu_j + (1-p)u_j = u_j.$$

1) Man verwechsle eine 'hypothetische' Entscheidungssituation nicht mit der tatsächlichen. In der ersten werden hypothetische Wahrscheinlichkeitsverteilungen über Konsequenzen als Lotterien zum heuristischen Zweck der Herleitung von Nutzwertigkeiten angenommen. (In der tatsächlichen Wahlsituation ist ein solches Verfahren unzulässig). Im obigen Beispiel sind somit $U_p, \bar{U}_p$ und H_1 hypothetische Umstände und hypothetische Handlungen.

Falls das Individuum in der obigen hypothetischen Situation $H_1(p)$ "wählt", dann bedeutet das $pu_1 + (1-p)u_3 \geq u_j$; falls es dagegen H_2 "wählt", dann ist $u_j \geq pu_1 + (1-p)u_r$.

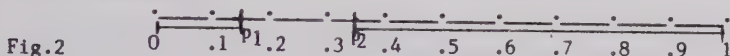
Daraus ergibt sich: Falls das Individuum $H_1(p)$ und nicht H_2 für den Umstand U_p , dessen Wahrscheinlichkeit p ist, "wählt", dann wird es $H_1(p')$ und nicht H_2 für jeden Umstand $U_{p'}$ mit $p' \geq p$ "wählen"; falls das Individuum H_2 und nicht $H_1(p)$ für den Umstand U_p , dessen Wahrscheinlichkeit p ist, "wählt", dann wird es H_2 und nicht $H_1(p')$ für jeden Umstand $U_{p'}$ mit $p' \leq p$ "wählen".

Mit anderen Worten, je größer p ist, desto größer wird auch die Attraktivität oder Wünschbarkeit von $H_1(p)$ für das Individuum im Vergleich zu der "sure-thing"-Handlung H_2 sein.¹⁾

Empirische Messung:

Das Individuum wird aufgefordert, den größten Wert von p anzugeben, bei dem $H_2 \succ H_1(p)$ gilt. (Dieser Wert ist dann α_j); sodann den kleinsten Wert von p , für den $H_1(p) \succ H_2$ gilt (dieser Wert ist dann β_j).

Anstatt einer linearen Nützlichkeitskala benutzen wir jetzt eine Wahrscheinlichkeitsskala



Das Individuum muß zwei Intervalle angeben:

1. das Intervall von 0 nach rechts mit $R_j \succ (R_1, p, R_r)$ für alle p ;

2. das Intervall von 1 nach links mit $(R_1, p, R_r) \succ R_j$ für alle p .

α_j ist dann p_1 und β_j p_2 . Der Abstand zwischen diesen beiden Intervallen soll möglichst gering sein.

(4) Beschränkte Intervallmetrisierung zweiter Stufe

Diese Methode besteht im Bestimmen des Verhältnisses zwischen zwei Wertintervallen $(u_j - u_{j+1})$ und $(u_1 - u_r)$ für $j = 1, \dots, r-1$.

Die Ordnung $u_1 \geq u_2 \geq \dots \geq u_r$ mit $u_1 > u_r$ sei gegeben.

Sei

$$L_j = \frac{u_j - u_{j+1}}{u_1 - u_r}$$

weiter seien α_j und $\alpha_j + \epsilon_j$, $0 \leq \alpha_j \leq \alpha_j + \epsilon_j \leq 1$, die Schranken von L_j .

Es gilt

$$\alpha_j \leq L_j \leq \alpha_j + \epsilon_j, \quad j = 1, \dots, r-1,$$

$$\text{mit } \sum_{j=1}^{r-1} \alpha_j \leq 1 \leq \sum_{j=1}^{r-1} \alpha_j + \sum_{j=1}^{r-1} \epsilon_j \quad (\text{wegen } \sum_{j=1}^{r-1} L_j = 1)$$

Wir wollen die Schranken von L_j mit Hilfe hypothetischer Lotterien bestimmen.

Es gilt:

1) Vgl. Fishburn, S.102.

$$L_1 = \frac{u_1 - u_2}{u_1 - u_r} = \frac{u_1 - u_r}{u_1 - u_r} - \frac{u_2 - u_r}{u_1 - u_r} = 1 - I_2,$$

$$L_{r-1} = \frac{u_{r-1} - u_r}{u_1 - u_r} = I_{r-1}.$$

Die Schranken von L_1 und L_{r-1} lassen sich also aufgrund der vorangegangenen Methoden aus (3) bestimmen.

Für L_j , $2 \leq j \leq r-2$, ergibt sich die Nutzenordnung

$$u_1 \geq u_j \geq u_{j+1} \geq u_r$$

aus der Präferenzordnung

$$R_1 \succ R_j \succ R_{j+1} \succ R_r.$$

Man betrachte nun die beiden Lotterien

$$(R_1, p, R_{j+1}) \text{ und } (R_r, p, R_j) \quad 0 \leq p \leq 1,$$

die durch die folgende hypothetische Situation veranschaulicht werden können

	U_p	$\bar{U}_p$
$H_1(p)$	R_1	R_{j+1}
$H_2(p)$	R_r	R_j

Falls H_1 "gewählt" wird, dann "ereignet sich" R_1 mit der Wahrscheinlichkeit p und R_{j+1} mit der Wahrscheinlichkeit $1-p$. Falls H_2 "gewählt" wird, dann "ereignet sich" R_r mit der Wahrscheinlichkeit p und R_j mit der Wahrscheinlichkeit $1-p$.

Man betrachte als Beispiel die folgende konkrete Situation (mit Geldpreisen als Konsequenzen):

	U_p	$\bar{U}_p$
$H_1(p)$	\$100	\$60
$H_2(p)$	\$1	\$80

Für $p = 1/2$ ergibt sich eine symmetrische Situation (in diesem Fall liefert H_1 \$100 bzw. \$60 jeweils mit $p = 1/2$; dasselbe gilt für H_2 in bezug auf \$80 und \$1) und somit die Gleichheit

	$U_{1/2}$	$\bar{U}_{1/2}$		$U_{1/2}$	$\bar{U}_{1/2}$	
$H_1(1/2)$	\$100	\$60	=	$H_1(1/2)$	\$100	\$60
$H_2(1/2)$	\$80	\$1		$H_2(1/2)$	\$1	\$80

bzw. allgemein

	$U_{1/2}$	$\bar{U}_{1/2}$		$U_{1/2}$	$\bar{U}_{1/2}$	
$H_1(1/2)$	R_1	R_{j+1}	=	$H_1(1/2)$	R_1	R_{j+1}
$H_2(1/2)$	R_j	R_r		$H_2(1/2)$	R_r	R_j

Da $R_1 \succ R_j$ und $R_{j+1} \succ R_r$, ist es plausibel, daß $H_1(1/2)$ und nicht $H_2(1/2)$ für $p = 1/2$ "gewählt" wird.

Mit diesen Bemerkungen sind wir in der Lage, die Axiome für die beschränkten Intervallmetrisierungen zweiter Stufe anzugeben.

Neben A1 und A3(2) sind es:

- A6(1) Für alle $j, j=1, \dots, r-1$ gibt es einen Wert $p \geq 0$, etwa p_j , mit $(R_r, p, R_j) \succ (R_1, p, R_{j+1})$ für alle $p \leq p_j$.
- A6(2) Für alle $j, j=1, \dots, r-1$, gibt es einen Wert von p , etwa q_j , mit $p_j \leq q_j \leq 1/2$ und $(R_1, p, R_{j+1}) \succ (R_r, p, R_j)$ für alle $p \geq q_j$.
- A6(3) Wenn $(R_1, p, R_{j+1}) \succ (R_r, p, R_j)$, dann $pu_1 + (1-p)u_{j+1} \geq pu_r + (1-p)u_j$; und wenn $(R_r, p, R_j) \succ (R_1, p, R_{j+1})$, dann $pu_r + (1-p)u_j \geq pu_1 + (1-p)u_{j+1}$.

Daraus folgt einerseits

$$\frac{p}{1-p} \leq L_j,$$

insbesondere

$$\frac{p_j}{1-p_j} \leq L_j;$$

andererseits

$$L_j \leq \frac{p}{1-p},$$

insbesondere

$$L_j \leq \frac{q_j}{1-q_j}.$$

Kombiniert ergibt dies

$$\alpha_j \leq L_j \leq \alpha_j + \epsilon_j, \quad j = 1, \dots, r-1,$$

$$\alpha_j = \frac{p_j}{1-p_j} \quad \alpha_j + \epsilon_j = \frac{q_j}{1-q_j}$$

für $0 \leq p_j \leq q_j \leq 1/2$ gilt $0 \leq \alpha_j \leq \alpha_j + \epsilon_j \leq 1$.

$$A6(4) \quad \sum_{j=1}^{r-1} \frac{p_j}{1-p_j} \leq 1 \leq \sum_{j=1}^{r-1} \frac{q_j}{1-q_j}.$$

Das garantiert:

$$\sum \alpha_j \leq 1 \leq \sum \alpha_j + \sum \epsilon_j.$$

Bei Anwendung dieser Methode muß das Individuum p_j möglichst groß wählen mit $(R_r, p_j, R_j) \succ (R_1, p_j, R_{j+1})$ und q_j möglichst klein mit $(R_1, q_j, R_{j+1}) \succ (R_r, q_j, R_j)$.

(5) Beschränkte Intervallmetrisierung dritter Stufe

Es gelten A1(1), A1(2) und A3.

In der Menge der Konsequenzen $\mathcal{R}$ betrachten wir die Teilmenge $\mathcal{R}'$, für deren Elemente die strikte Präferenzordnung, d.h. $R_1 \succ R_2 \succ \dots \succ R_{r'}$, gilt. Aus dieser Präferenzordnung ist die Nutzenordnung, die auch strikt ist, ableitbar (nach A1), d.h.

$$u_1 > u_2 > \dots > u_{r'}, \quad r' \leq r.$$

(die relativen Werte aller anderen Konsequenzen, die nicht in der ausgewählten Menge sind, werden mit einem der u_j , $j = 1, \dots, r'$, identisch sein).

Die Metrisierung besteht darin, daß man das Intervall bestimmt, in welches der Quotient aus den Wertintervallen $(u_j - u_{j+1})$ und $(u_{j+1} - u_{j+2})$, $j = 1, \dots, r'-2$, fällt.

Sei

$$K_j = \frac{u_j - u_{j+1}}{u_{j+1} - u_{j+2}} \quad j = 1, 2, \dots, r'-2.$$

Weiter seien c_j und d_j Schranken von K_j mit:

$$\begin{aligned} c_j &\leq K_j \leq d_j, & j &= 1, \dots, r'-2 \\ 0 &\leq c_j \leq d_j \leq \infty & j &= 1, \dots, r'-2. \end{aligned}$$

Die Schranken von K_j können wie unter (3) bestimmt werden.

1.Methode:

Das Individuum gibt auf einer linearen Skala mit den Endpunkten u_j und u_{j+2} ein Intervall an, welches seiner Meinung nach u_{j+1} enthält:

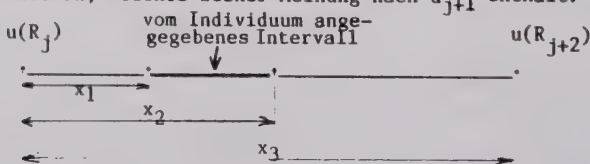


Fig.3

Die Schranken von K_j sind gegeben durch

$$c_j = \frac{x_1}{x_3 - x_1} \quad d_j = \frac{x_2}{x_3 - x_2},$$

x_1, x_2, x_3 sind lineare Abstände (wie auf Fig.3); wir nehmen an, daß $x_2 < x_3$ ist.

2.Methode:

Man vergleicht R_{j+1} mit (R_j, p, R_{j+2}) . (R_j, p, R_{j+2}) ist dabei die hypothetische Lotterie (vgl.S. 411 dieser Arbeit), die zu R_j mit der Wahrscheinlichkeit p oder zu R_{j+2} mit der Wahrscheinlichkeit $1-p$ "führt". Dabei gelten (A5 ähnliche) Axiome:

- A7(1) Für alle j , $j = 1, \dots, r-2$, gibt es einen Wert von p , etwa p_j , mit $R_{j+1} \succ (R_j, p, R_{j+2})$ für alle $p \leq p_j$.
- A7(2) Es gibt einen Wert von p , etwa q_j , mit $q_j \geq p_j$, und $(R_j, p, R_{j+2}) \succ R_{j+1}$ für alle $p \geq q_j$.
- A7(3) Wenn $R_{j+1} \succ (R_j, p, R_{j+2})$, dann $u_{j+1} \geq pu_j + (1-p)u_{j+2}$; und wenn $(R_j, p, R_{j+2}) \succ R_{j+1}$, dann $pu_j + (1-p)u_{j+2} \geq u_{j+1}$.

Aus den Axiomen folgt
einerseits

$$\frac{1-p}{p} \geq \frac{u_j - u_{j+1}}{u_{j+1} - u_{j+2}} = K_j, \quad K_j \leq \frac{1-p}{p},$$

insbesondere

$$K_j \leq \frac{1-p_j}{p_j};$$

andererseits

$$\frac{1-p}{p} \leq K_j,$$

insbesondere

$$K_j \geq \frac{1-q_j}{q_j}.$$

Daraus ergibt sich

$$c_j \leq K_j \leq d_j, \quad c_j = \frac{1-q_j}{q_j}, \quad d_j = \frac{1-p_j}{p_j}, \quad j = 1, \dots, r'-2.$$

(6) geordnete Metrisierungen (Metrisierung erster Ordnung, Metrisierung n -ter Ordnung, $n = 2, \dots$)

Die Metrisierung erster Ordnung basiert auf der Nutzenordnung $u_1 \geq u_2 \geq \dots \geq u_r$ und die Metrisierung n -ter Ordnung auf der $(n-1)$ -ten metrisierten Ordnung. Die Exaktheit bei der Bestimmung von Nützlichkeitswerten vergrößert sich mit der Höhe der Ordnung.

Die Metrisierung erster Ordnung besteht darin¹⁾ daß man die $r-1$ Differenzen von u_j der Größe nach ordnet. Diese Differenzen sind $u_1 - u_2, u_2 - u_3, \dots, u_{r-1} - u_r$. Mit $j_1, j_2, \dots, j_{r-1}$ als Permutationen von $1, 2, \dots, r-1$ hat diese Ordnung die Form

1) In folgendem nehmen wir an, daß auf jeder Menge, die aus Nützlichkeitsdifferenzen besteht, eine schwache Ordnung ableitbar ist, d.h. daß neben A1 (das für die zugrundeliegende Nutzenordnung der relativen Werte gilt) auf jeder Menge, die aus Nützlichkeitsdifferenzen besteht, ein ähnliches Axiom gilt, das die schwache Ordnung auf dieser Menge garantiert.

$$u_{j_1} - u_{j_1+1} \geq u_{j_2} - u_{j_2+1} \geq \dots \geq u_{j_{r-1}} - u_{j_{r-1}+1} \geq 0,$$

bzw. mit $u_k^{(1)} = u_{j_k} - u_{j_k+1}$, $k = 1, \dots, r-1$, $u_r^{(1)} = 0$,

$$u_1^{(1)} \geq u_2^{(1)} \geq \dots \geq u_{r-1}^{(1)} \geq u_r^{(1)}.$$

Die Metrisierung zweiter Ordnung besteht darin, daß man die Differenzen $u_j^{(1)} - u_{j+1}^{(1)}$, $j = 1, \dots, r-1$, d.h. die Differenzen der Nützlichkeitswerte, die durch die vorangegangene Metrisierung (erster Ordnung) entstanden sind, der Größe nach ordnet.

Sei $u_k^{(2)}$ die k -te Differenz $u_j^{(1)} - u_{j+1}^{(1)}$, und $u_r^{(2)} = 0$. Diese Ordnung hat dann die Gestalt

$$u_1^{(2)} \geq u_2^{(2)} \geq \dots \geq u_{r-1}^{(2)} \geq u_r^{(2)}.$$

Gegeben sei die $(n-1)$ -te Differenzordnung

$$u_1^{(n-1)} \geq u_2^{(n-1)} \geq \dots \geq u_{r-1}^{(n-1)} \geq u_r^{(n-1)}.$$

Auf analoge Weise wie oben erhält man die n -te Ordnung:

$$u_1^{(n)} \geq u_2^{(n)} \geq \dots \geq u_{r-1}^{(n)} \geq u_r^{(n)}.$$

Die sukzessiv fortschreitenden Metrisierungen von Nutzendifferenzen ergeben immer feinere Grenzen für u_j .

Beispiel:

Sei die folgende Präferenzordnung

$$R_1 \succ R_2 \succ R_3 \succ R_4$$

gegeben. Daraus ergibt sich nach A1 die Nutzenordnung

$$u_1 \geq u_2 \geq u_3 \geq u_4.$$

Wir ordnen den Konsequenzen R_1 und R_4 willkürliche Nützlichkeitswerte zu, etwa $u(R_1) = 4$ und $u(R_4) = 0$.

Zur Bestimmung von $u(R_2)$ und $u(R_3)$ machen wir folgende Annahmen:

$$(*) \quad u_3 - u_4 > u_1 - u_2 = u_2 - u_3 > u_4 - u_4 = 0$$

(Metrisierung erster Ordnung).

Mit $u_1 = 4$ und $u_4 = 0$ sieht die obige Nutzenordnung wie folgt aus

$$u_3 > 4 - u_2 = u_2 - u_3 > 0.$$

Eine Lösung ist $u(R_2) = 3$ und $u(R_3) = 2$.

Angenommen, in $(*)$ (mit Abk. $r_1 = u_3 - u_4$, $r_2 = u_1 - u_2$, $r_3 = u_2 - u_3$, $r_4 = u_4 - u_4$) gilt

$$r_1 - r_2 = r_3 - r_4 > r_2 - r_3 = r_4 - r_4$$

(Metrisierung zweiter Ordnung),

d.h. es gilt

$$u_3 - u_4 - u_1 + u_2 = u_2 - u_3 - u_4 + u_4 > u_1 - u_2 - u_2 + u_3 = u_4 - u_4 - u_4 + u_4$$

oder

$$u_3 - 4 + u_2 = u_2 - u_3 > 4 - 2u_2 + u_3 = 0$$

oder

$$u_3 - 4 = -u_3 > 4 - 3u_2 + u_3 = -u_2.$$

Daraus folgt die exakte Lösung:

$$u(R_2) = 3 \quad \text{und} \quad u(R_3) = 2.$$

(7) Metrisierung höherer Ordnung

Im Unterschied zur Metrisierung n -ter Ordnung ($n \geq 1$) besteht die Metrisierung höherer Ordnung darin, daß man alle Differenzen $u_j - u_k$, $j < k$, der Größe nach ordnet. (Bei der Metrisierung n -ter Ordnung ($n \geq 1$) hatten wir es nur mit den Differenzen der Form $u_j - u_{j+1}$, $j = 1, \dots, r-1$, zu tun).

Sei $u_1 \geq u_2 \geq \dots \geq u_r$ die Nutzenordnung. Mit $r = 4$ etwa erhalten wir folgende Nutzendifferenzen:

$$u_1 - u_2, u_1 - u_3, u_1 - u_4, u_2 - u_3, u_2 - u_4, u_3 - u_4.^{1)}$$

Beispiel: wie vorher

Diesmal nehmen wir jedoch an, daß die folgende Nutzenordnung ableitbar ist

$$u_1 - u_4 > u_2 - u_4 > u_1 - u_3 = u_3 - u_4 > u_1 - u_2 = u_2 - u_3,$$

oder (mit $u(R_1) = 4$ und $u(R_4) = 0$ wie vorher)

$$4 > u_2 > 4 - u_3 = u_3 > 4 - u_2 = u_2 - u_3.$$

Aus $4 - u_3 = u_3$ folgt $u_3 = 2$, und aus $4 - u_2 = u_2 - u_3$ ergibt sich $u_2 = 3$.

Somit ist die Lösung eindeutig: $u(R_2) = 3$ und $u(R_3) = 2$.

(8) Intervallmetrisierungen

Es handelt sich hier um die exakte Methode zur Bestimmung von Nützlichkeitswerten. Es gelten dabei Axiome, wie sie etwa von v. Neumann und Morgenstern (1944)²⁾ oder Savage (1954)³⁾ aufgestellt wurden. Diese Axiome garantieren die Existenz und Eindeutigkeit bis auf positiv lineare Transformationen einer Nützlichkeitsfunktion. Den v. Neumann-Morgenstern Axiomen liegt der Begriff des "Glücksspiels" oder der "Lotterie" (vgl. S. 411 dieser Arbeit) zugrunde. Ihr Wahrscheinlichkeitsbegriff ist objektiv und daher weniger anwendungsfähig. Savage, dem diese praktische Schwierigkeit bewußt war, führte den Begriff der personellen (oder subjektiven) Wahrscheinlichkeit in sein axiomatisches System ein. (Die v. Neumann-Morgenstern-Axiome wie auch diejenigen von Savage findet der Leser im Anhang II. u. III).

1) für r beliebig gibt es insgesamt $\frac{r(r-1)}{2}$ Nutzendifferenzen.

2) v. Neumann, J.-Morgenstern, O., Theory of Games and Economic Behavior (1944).

3) Savage, L. J., The Foundations of Statistics (1954).

Wir beginnen damit, daß wir mit Hilfe von A3(1) und A3(2) bestimmen, welche Konsequenzen in der (endlichen) Menge der Konsequenzen $\mathcal{R}$ am wünschenswertesten und welche am wenigsten wünschenswert sind. Seien diese etwa R_1 und R_r ; sodann (vgl. Luce und Raiffa, S.30¹⁾) versuchen wir, einen Wahrscheinlichkeitswert von p , etwa p_j , herauszufinden, bei dem das Individuum in bezug auf R_j und die Lotterie (R_1, p_j, R_r) indifferent ist. Diese Prozedur wiederholen wir für jedes j , $j = 2, \dots, r-1$.

Es gilt das folgende Axiom:

A8 Es seien $R_1 \succ R_j \succ R_r$ und $R_1 \succ R_r$ gegeben; für jedes j , $j = 2, \dots, r-1$, existiert eine Wahrscheinlichkeit p_j , $0 \leq p_j \leq 1$, so daß $R_j \sim (R_1, p_j, R_r)$ und $(R_1, p_j, R_r) \succ R_j$ (d.h. $R_j \sim (R_1, p_j, R_r)$ ist. Weiter gilt: $R_j \succ (R_1, p, R_r)$ für jedes $p < p_j$ und $(R_1, p, R_r) \succ R_j$ für jedes $p > p_j$. Außerdem gilt: $u(R_j) \geq pu(R_1) + (1-p)u(R_r)$ genau dann, wenn $R_j \succ (R_1, p, R_r)$.

Wir setzen $u(R_1) = 1$ und $u(R_r) = 0$. Daraus ergibt sich $u(R_j) = p_j$, $j = 2, \dots, r-1$, bzw. allgemeiner

$$I_j = \frac{u_j - u_r}{u_1 - u_r} = p_j$$

oder

$$u_j = p_j(u_1 - u_r) + u_r.$$

Beispiel:

(Wir benutzen dabei den Begriff der hypothetischen Lotterie)

Angenommen, wir haben vier Konsequenzen: "Der Mörder wird bestraft" (R_2), "Der Nicht-Mörder wird bestraft" (R_4), "Der Mörder wird nicht bestraft" (R_3), "Der Nicht-Mörder wird nicht bestraft" (R_1).

Das Individuum entscheide sich für die folgende Präferenzordnung

$$R_1 \succ R_2 \succ R_3 \succ R_4$$

Für den Vergleich R_j mit (R_1, p, R_r) mit $j = 2, 3$ und $r = 4$, konstruieren wir folgende hypothetische Situation:

$$\begin{array}{cc} U & \bar{U} \\ H_1 & \left[\begin{array}{cc} R_1 & R_r \end{array} \right] \\ H_2 & \left[\begin{array}{cc} R_j & R_j \end{array} \right] \end{array} \quad j = 2, 3; r = 4.$$

Angenommen, das Individuum ist der Ansicht, daß

$$R_2 \sim (R_1, 9/10, R_4) \text{ und } R_3 \sim (R_1, 1/2, R_4).$$

Daraus folgt nach A8

$$u(R_2) = 9/10 \cdot u(R_1) + 1/10 \cdot u(R_4)$$

$$u(R_3) = 1/2 \cdot u(R_1) + 1/2 \cdot u(R_4).$$

1) Luce & Raiffa benutzen in ihrem Buch, Games and Decisions (1957) Axiome, die denen von v. Neumann & Morgenstern ähnlich sind.

Mit $u(R_1) = 1$ und $u(R_4) = 0$ ergibt sich daraus

$$u(R_2) = 9/10 \quad \text{und} \quad u(R_3) = 1/2.$$

Wir haben somit (aufgrund der Präferenzordnung des Individuums) die folgende Nützlichkeitsmatrix abgeleitet:

	Mord	kein Mord
Strafe	$\left[\begin{array}{c} 9/10 \end{array} \right]$	$\left[\begin{array}{c} 0 \end{array} \right]$
keine Strafe	$\left[\begin{array}{c} 1/2 \end{array} \right]$	$\left[\begin{array}{c} 1 \end{array} \right]$

Angenommen, das Individuum ist imstande, die Werte von p_2 bis p_{r-1} in der vom Axiom A8 postulierten Weise anzugeben, wobei es von den folgenden Ordnungen ausgeht

$$R_1 \succ R_2 \succ \dots \succ R_r \quad \text{und} \quad p_1 \geq p_2 \geq \dots \geq p_r \quad \text{mit} \quad p_1 = 1, \quad p_r = 0.$$

(Man vergleicht stets R_j mit (R_1, p_j, R_r) , $j = 1, \dots, r$).

Wenn das Individuum imstande ist, den Wert von p anzugeben mit $R_j \sim (R_1, p, R_r)$, dann könnte man auch erwarten, daß es den Wert von p angeben kann mit $R_j \sim (R_i, p, R_k)$ für $i < j < k$. Sei q_{ijk} der Wert von p , für welches $R_j \sim (R_i, p, R_k)$ gilt. Dann ist $R_j \sim (R_i, q_{ijk}, R_k)$. Daher könnte man erwarten, daß auch

$$u_j = q_{ijk}u_i + (1 - q_{ijk})u_k$$

gilt.

Nach den v. Neuman & Morgenstern oder Luce & Raiffa Axiomen ist das auch der Fall. Seien $u_1 = 1$ und $u_r = 0$; die obige Gleichung lautet dann umgeschrieben:

$$p_j = q_{ijk}p_i + (1 - q_{ijk})p_k \quad i < j < k.$$

Man kann die letzte Gleichung dazu benutzen, um p_j und q_{ijk} sukzessiv zu verändern, bis diese Gleichung für alle Fälle erfüllt ist.

I.2. Metrisierung von Wahrscheinlichkeiten

(1) Qualitative Metrisierung

Wir benutzen folgende Bezeichnungen:

$\mathcal{U} = \{U_1, \dots, U_n\}$ ist eine endliche Menge, deren Elemente Umstände U_i , $i = 1, \dots, n$, sind;

$\succsim$ ist eine binäre Relation (zu lesen: "nicht weniger wahrscheinlich als"), die auf der Menge $\mathcal{U}$ definiert ist.

Es gilt das folgende Axiom:

Ordnungsaxiom: $\succsim$ ist eine schwache Ordnung auf der Menge $\mathcal{U}$.

Aufgrund dieses Axioms ergibt sich eine Ordnung der U_i

$$U_1 \succsim U_2 \succsim U_3 \succsim \dots \succsim U_n \succsim \emptyset^1)$$

Dieser Ordnung korrespondiert die schwache Ordnung der Wahrscheinlichkeiten der Umstände.

$$P(U_1) \geq P(U_2) \geq \dots \geq P(U_n) \geq 0, \quad \sum_{i=1}^n P(U_i) = 1.$$

Die schwache Ordnung $U_1 \succsim U_2 \succsim \dots \succsim U_n$ kann auf dieselbe Weise gewonnen werden wie die schwache (Präferenz-) Ordnung der Konsequenzen (ordinale Metrisierung, S.408).

Um U_j mit U_k zu vergleichen, können wir auch den entscheidungstheoretischen Ansatz benutzen (vgl. S.411f).

Betrachten wir nun zwei hypothetische Handlungen²⁾ H_1 und H_2 mit $H_1(U_j) = h$, $H_1(U_i) = h'$, $i \neq j$, $H_2(U_k) = h$, $H_2(U_i) = h'$, $i \neq k$, wobei h und h' zwei hypothetische Konsequenzen sind mit $h \succ h'$. Wir wollen nun feststellen, ob $U_j \succsim U_k$ oder $U_k \succsim U_j$. Es gilt für alle Paare h, h' mit $h \succ h'$ entweder $H_1 \succsim H_2$ oder $H_2 \succsim H_1$. Man definiert: $U_j \succsim U_k$ genau dann, wenn $H_1 \succsim H_2$; und $U_k \succsim U_j$ genau dann, wenn $H_2 \succsim H_1$. (vgl. dazu die ersten vier Postulate von Savage, S.431f).

Beispiel:

Es seien drei Umstände: "fahrlässige Tötung" (U_1), "Totschlag" (U_2) und "Mord" (U_3) und zwei Handlungen: "Strafe₁" (H_1) und "Strafe₂" (H_2) gegeben.

Wir betrachten die folgenden Konsequenzen: $(H_1, U_1) = \alpha$, $(H_1, U_2) = \alpha'$, $(H_2, U_3) = \alpha$ und $(H_2, U_2) = \alpha'$. Außerdem gelte $\alpha \succ \alpha'$. Angenommen, das Individuum ist der

1) $\emptyset \hat{=}$ Nullereignis bzw. das logisch unmögliche Ereignis.

2) Hypothetisch ist die o.Situation, weil es nicht um die tatsächliche Wahl einer Handlung geht, sondern um die Bestimmung von Wahrscheinlichkeitspräferenzen vgl. dazu S.411f, wo wir den Hilfsbegriff "hypothetische Entscheidungssituation" bei der Herleitung von Nützlichkeitswerten benutzt haben.

Ansicht, daß in dieser Situation H_1 (hypothetisch) H_2 vorzuziehen sei. Nach den obigen Erläuterungen bedeutet das nichts anderes, als daß das Individuum U_1 für wahrscheinlicher hält als U_3 .

Den Vergleich zwischen U_j können wir auch auf eine andere Weise vornehmen. Man betrachtet anstatt Umständen die Mengen von Umständen bzw. Ereignisse. n Umstände ergeben 2^n Ereignisse (Nullmenge und Gesamtmenge inbegriffen). So ergeben drei Umstände im vorigen Beispiel 8 Ereignisse:

$$\begin{array}{ll} E_1 = \{\emptyset\} & E_5 = \{U_1, U_2\} \\ E_2 = \{U_1\} & E_6 = \{U_1, U_3\} \\ E_3 = \{U_2\} & E_7 = \{U_2, U_3\} \\ E_4 = \{U_3\} & E_8 = \{U_1, U_2, U_3\} \end{array}$$

$\{U_i, U_j\}$ bedeutet ein Ereignis, das aus den Umständen U_i und U_j besteht; $P(U_i \vee U_j) = P(U_i) + P(U_j)$, die Wahrscheinlichkeit, daß entweder U_i oder U_j sich ereignet ($U_i \wedge U_j = \emptyset$).

Die Ereignisse E_i , $i = 1, \dots, 8$, können schwach geordnet werden.

Angenommen, durch paarweisen Vergleich haben wir herausgefunden

$$E_3 \succcurlyeq E_4 \quad \text{und} \quad E_2 \succcurlyeq E_7.$$

Wir erhalten somit

$$E_8 \succcurlyeq E_5 \succcurlyeq E_6 \succcurlyeq E_2 \succcurlyeq E_7 \succcurlyeq E_3 \succcurlyeq E_4 \succcurlyeq E_1$$

oder

$$1 \geq p_1 + p_2 \geq p_1 + p_3 \geq p_1 \geq p_2 + p_3 \geq p_2 \geq p_3 \geq 0,$$

diese Ordnung ist vollständig bestimmt durch

$$p_2 \geq p_3 \quad \text{und} \quad p_1 \geq p_2 + p_3 \quad \text{mit} \quad p_j \geq 0 \quad \text{und} \quad \sum_{j=1}^n p_j = 1.$$

(für weitere Vergleichsmethoden siehe Fishburn, S.188f).

(2) Intervallmetrisierung von subjektiven Wahrscheinlichkeiten

Unter einer Intervallmetrisierung von subjektiven Wahrscheinlichkeiten verstehen wir eine Metrisierung, die jedes p_j auf ein Intervall, das im Intervall $[0,1]$ enthalten ist, einschränkt. Seien E_j , $j = 1, \dots, r$, im folgenden Ereignisse.

Weiter sei $p_j = P(E_j)$ mit $\sum_{j=1}^n p_j = 1$.

Eine beschränkte Intervallmetrisierung von p_j hat die Form

$$\alpha_j \leq p_j \leq \beta_j, \quad j = 1, \dots, r,$$

mit $[\alpha_j, \beta_j] \subset [0,1]$, d.h. $0 \leq \alpha_j \leq \beta_j \leq 1$

und $\sum_{j=1}^r \alpha_j \leq 1 \leq \sum_{j=1}^r \beta_j$.

Wenn $\alpha_j = \beta_j$, dann erhalten wir den exakten Wert von p_j .

Mit $\varepsilon_j = \beta_j - \alpha_j$ sehen die obigen Ungleichungen so aus

$$\alpha_j \leq p_j \leq \alpha_j + \varepsilon_j$$

(diese Schreibweise gibt einen Unsicherheitsraum des Individuums bei der Bestimmung von p_j an).

Eine einfache (intuitive) Methode, die Schranken von p_j zu bestimmen, besteht in der Befragung des Individuums. Das Individuum wird aufgefordert, die Zahlen α_j und β_j , $0 \leq \alpha_j, \beta_j \leq 1$ anzugeben, die seiner Meinung nach p_j mit Sicherheit enthalten.

Eine andere Methode ist die von Schlaifer (1959). Der Grundbegriff ist hier der der hypothetischen "standard lottery" (vgl. Fishburn, S.179). Diese Methode soll hier kurz an einem Beispiel geschildert werden.

Wir nehmen an, daß die Standardlotterie aus N möglichen Resultaten besteht, die vom Individuum für gleich wahrscheinlich gehalten werden. Die "Grundhandlung" in einer solchen Lotterie bestehe etwa in der Ziehung einer Kugel aus einer Urne, die die von 1 bis N durchnummerierten Kugeln enthält.¹⁾ Weiter sei C_n das Ereignis "Eine von $n \leq N$ spezifizierten Kugeln wird gezogen". E_j und $\neg E_j$ seien Ereignisse, an deren Realisierung wir interessiert sind: E_j führe zu einer hypothetischen Konsequenz h_1 und $\neg E_j$ zu einer hypothetischen Konsequenz h_2 mit $h_1 \succ h_2$. Analog führe C_n zu h_1 und $\neg C_n$ zu h_2 (d.h. während die Standardlotterie eine völlig hypothetische Situation - was Ereignisse, Konsequenzen und Handlungen angeht - beschreibt, ist die E_j -Situation sozusagen "pseudohypothetisch": Die Umstände darin sind aktuelle Umstände der realen Welt, erst die Handlungen und die Konsequenzen sind hier hypothetisch)²⁾. Nun versuchen wir, zwei Werte von n , etwa n_1 und n_2 , $n_1 \leq n_2$, herauszufinden, so daß gilt:

1. Für alle $n \leq n_1$ zieht das Individuum die E_j -Situation der Standardlotterie mit C_n vor;
2. Für alle $n \geq n_2$ zieht das Individuum die Standardlotterie mit C_n der E_j -Situation vor³⁾.

Die Schranken von $P(E_j)$ sind dann

$$\frac{n_1}{N} \leq P(E_j) \leq \frac{n_2}{N}.$$

1) Erläuterung: Sei $n = 5$. Das bedeutet, das Individuum muß aus der Urne fünf Kugeln ziehen (mit Zurücklegen nach jeder Ziehung). Diese Kugeln müssen unterschiedlichen sein (d.h. evtl. nachziehen, wenn ein und dieselbe Zahl mehrmals gezogen wird); dann zieht das Individuum aus der Urne erneut eine Kugel. Falls diese eine der gezogenen Zahlen trägt, dann ist das Ereignis C_5 realisiert, sonst $\neg C_5$.

2) Es gilt $P(C_n) = \frac{n}{N}$, wobei N die Gesamtheit der Kugeln und n die Anzahl der spezifizierten Kugeln ist. Ebenso ist $P(E_j) = \frac{n}{N}$.

3) Eine andere Methode besteht darin, einen Wert von n direkt anzugeben, bei dem das Individuum zwischen der hypothetischen Standardlotterie und der E_j -Situation indifferent ist. (vgl. Fishburn, S.179). Falls das Individuum in bezug auf die beiden hypothetischen Handlungen, die jeweils h_1 zu E_j bzw. C_n (die erste Handlung) und h_2 zu $\neg E_j$ bzw. $\neg C_n$ (die zweite Handlung) zuordnen, bei einem bestimmten n indifferent ist, dann ist

$$P(C_n) = \frac{n}{N} \text{ und ebenfalls } P(E_j) = \frac{n}{N}.$$

Es ist wünschenswert, daß n_1 und n_2 möglichst nah beieinander liegen.

Man kann die Schranken von p_j wie folgt weiter verfeinern:

Sei $c = 1 - \sum \alpha_j \geq 0$.

Man ersetze β_j durch $\beta'_j = \alpha_j + c$, falls $\epsilon_j = \beta_j - \alpha_j > c$ für $j \in \{1, \dots, r\}$. Oder anders ausgedrückt, man ersetze β_k durch $\beta'_k = 1 - \sum_{j \neq k} \alpha_j$, falls $\sum_{j \neq k} \alpha_j + \beta_k < 1$ ist, und ersetze analog α_k durch $\alpha'_k = 1 - \sum_{j \neq k} \beta_j$, falls $\alpha_k + \sum_{j \neq k} \beta_j < 1$.

Zur Veranschaulichung: Sei $r = 4$. Die Anfangsschranken seien gegeben durch:

$$0.06 \leq p_1 \leq 0.2, \quad 0.16 \leq p_2 \leq 0.27, \quad 0.25 \leq p_3 \leq 0.4, \quad 0.05 \leq p_4 \leq 0.25.$$

Da $\beta_1 + \beta_2 + \beta_3 + \alpha_4 = 0.92$, können wir α_4 erhöhen von 0.05 auf 0.13.

Analog können wir, da $\alpha_1 + \beta_2 + \beta_3 + \beta_4 = 0.98$, α_1 erhöhen von 0.06 auf 0.08.

Eine weitere Methode, zu numerischen Wahrscheinlichkeiten zu gelangen, ergibt sich aufgrund der bereits abgeleiteten Nützlichkeiten und dem Axiom A8. Die Nützlichkeiten können wir dabei durch die ordinale Metrisierung oder durch die Metrisierung höherer Ordnung gewinnen. (Der Vorteil dieser beiden Methoden besteht gerade darin, daß sie von keinen Wahrscheinlichkeitsannahmen – auch nicht von solchen hypothetischer Art bei der Herleitung von Nützlichkeiten Gebrauch machen)

Im Beispiel auf S.413 haben wir die Nützlichkeitswerte: $u(R_1) = 4$, $u(R_2) = 3$, $u(R_3) = 2$ und $u(R_4) = 0$ hergeleitet. Aus A8 folgt, daß eine Wahrscheinlichkeit p_j , $0 \leq p_j \leq 1$, existiert, so daß $R_j \sim (R_1, p_j, R_r)$ bzw. $u(R_j) = p_j u(R_1) + (1-p_j)u(R_r)$ ist. Für alle $p < p_j$ ist $u(R_j) > pu(R_1) + (1-p)u(R_r)$ und für alle $p > p_j$ $pu(R_1) + (1-p)u(R_r) > u(R_j)$.

Falls das Individuum im obigen Beispiel zwischen R_2 und der hypothetischen Lotterie (R_1, p, R_4) indifferent ist (bzw. $u(R_2) = pu(R_1) + (1-p)u(R_4)$ setzt), so bedeutet dies nichts anderes, als das für dieses $p(U_1) = 3/4$ ist¹⁾. Für alle $p < 3/4$ muß daher $u(R_2) > pu(R_1) + (1-p)u(R_4)$ und für alle $p > 3/4$ $pu(R_1) + (1-p)u(R_4) > u(R_2)$ sein.

Natürlich kann das Individuum (R_1, p, R_4) auch mit R_3 vergleichen. Allerdings wird es dabei eine ganz andere Wahrscheinlichkeitsverteilung gewinnen, wenn es bezüglich (R_1, p, R_4) und R_3 indifferent ist. Sei im obigen Beispiel die folgende hypothetische Situation gegeben:

	U_1	U_2
H_1	4	0
H_2	3	2

Das Individuum sei der Ansicht, daß es (R_1, p, R_4) sowohl R_2 als auch R_3 vorziehe,

1) falls R_1 und R_r durch verschiedene Umstände U_1 und U_n mit $U_1 \neq U_n$ realisiert sind, so daß gilt $P(R_1) = P(U_1)$ und $P(R_r) = P(U_n)$.

d.h. für das Individuum ist $p \cdot 4 + (1-p) \cdot 0 > 3$ und $p \cdot 4 + (1-p) \cdot 0 > 2$. Daraus folgt, daß $p \cdot 4 > \max \{ u(R_2), u(R_3) \}$, also $p > 3/4$ sein muß.¹⁾

Eine andere Methode führt auch zum Ziel (auch dann, wenn ein und derselbe Umstand sowohl zu der wünschenswertesten als auch zu der unliebsamsten Konsequenz führt):

Sei die folgende Nützlichkeitsmatrix

	U_1	U_2
H_1	4	2
H_2	0	3

gegeben. Das Individuum betrachte nun die hypothetischen Erwartungswerte der beiden Handlungen, d.h.

$$E_U(H_1) = p \cdot 4 + (1-p) \cdot 2$$

$$E_U(H_2) = p \cdot 0 + (1-p) \cdot 3.$$

Nehmen wir weiter an, daß das Individuum in bezug auf die beiden Handlungen hypothetisch indifferent ist, dann kann es p wie folgt berechnen:

$$p \cdot 4 + (1-p) \cdot 2 = p \cdot 0 + (1-p) \cdot 3, \quad \text{also } p = 1/5.$$

Für alle $p > 1/5$ muß daher das Individuum H_1 und für alle $p < 1/5$ H_2 vorziehen (d.h. falls das Individuum in der obigen Situation in bezug auf beide Handlungen nicht indifferent ist, bzw. H_1 H_2 vorzieht, so muß $P(U_1)$ in diesem Fall größer als $1/5$ sein).

Die Methode von Raiffa²⁾ In Form eines Dialogs zwischen einem Entscheidungsanalytiker und seinem Patienten führt Raiffa eine andere Methode zur Ermittlung subjektiver numerischer Wahrscheinlichkeitswerte vor. Wir wollen den Gedanken-gang hier kurz skizzieren, (Da wir uns mit einer juristischen Materie befassen, werden wir der wahrscheinlichkeitstheoretischen Beurteilung eine juristische Sachverhaltshypothese zugrunde legen).

Angenommen, das Individuum muß eine Hypothese wahrscheinlichkeitstheoretisch beurteilen, die wie folgt lautet: "Der Angeklagte hat den Diebstahl begangen". "p" bezeichnet im folgenden den gesuchten Wahrscheinlichkeitswert.

Die Methode besteht darin, daß man das gesuchte Intervall sukzessive einschränkt, so daß am Ende eine ziemlich genaue Schätzung für p herauskommt.

Als erstes wird eine Zahl genannt, die ein wenig größer ist als 0, z.B. 0.10, und das Individuum gefragt, ob es glaubt, daß p kleiner als 0.10 oder größer als 0.10 ist. Angenommen, das Individuum entscheidet sich für "größer als 0.10".

Es wird eine zweite Zahl genannt, die ein wenig kleiner ist als 1, z.B. 0.90, und das Individuum wiederum gefragt, ob es glaubt, daß p kleiner als 0.90 oder größer als 0.90 ist. Das Individuum entscheidet sich vielleicht für "kleiner".

1) In der obigen hypothetischen Situation setzten wir voraus, daß das Individuum in der Weise konsistent ist, als es nicht gleichzeitig $(R_1, p, R_4) \sim R_2$ und $(R_1, p, R_4) \sim R_3$ annimmt.

2) H. Raiffa, Einführung in die Entscheidungstheorie (1969), S.195-203).

Weiter wird das Individuum aufgefordert, einen Wert zu nennen, bei dem es ihm sehr schwer fallen würde zu sagen, ob p kleiner oder größer ist als dieser Wert (d.h. einen Wert, bei dem es für das Individuum gleichwahrscheinlich ist, daß p ober- oder unterhalb dieses Wertes liegt). Angenommen, das Individuum entscheidet sich für 0.60.

Durch das letzte Verfahren wird das Intervall von 0 bis 1 durch 0.60 in zwei urteilsmäßig gleichwahrscheinliche Teile geteilt. Dieser Prozeß wird wiederholt, indem die beiden Intervalle in verschiedene Unterintervalle aufgeteilt werden. Das Individuum wird z.B. gefragt, ob es glaubt, daß es wahrscheinlicher ist, daß p kleiner als 0.20 ist oder zwischen 0.20 und 0.60 liegt. Angenommen, das Individuum entscheidet sich für das letztere. Auf die Frage, ob p zwischen 0 und 0.58 oder zwischen 0.58 und 0.60 liege, könnte das Individuum etwa das erste Intervall nennen. Das Individuum wird dann aufgefordert, eine Zahl zu nennen, von der es glaubt, daß es gleichwahrscheinlich ist, daß p zwischen 0 und dieser Zahl oder zwischen dieser Zahl und 0.60 liegt. Das Individuum könnte sich z.B. für 0.50 entscheiden.

Danach wird das Individuum aufgefordert, das Intervall von 0.60 bis 1.00 in zwei gleichwahrscheinliche Teile zu teilen. Angenommen, das Individuum nennt 0.68. (Die vom Individuum oben genannten Zahlen, 0.60, 0.50, 0.68 drücken nichts anderes aus als seinen Glauben, daß es gleichwahrscheinlich ist, daß p in einem der vier Intervalle (0,0.50), (0.50,0.60), (0.60,0.68) und (0.68,1.00) liegt).

Man bezeichne die Zahl 0.60 als die subjektive 0.50-Fraktile des Individuums, die Zahl 0.50 als die subjektive 0.25-Fraktile und die Zahl 0.68 als die subjektive 0.75-Fraktile. Symbolisch wird dies in der Form

$$p_{0.25} = 0.50; p_{0.50} = 0.60; p_{0.75} = 0.68$$

dargestellt.

Das Verfahren wird weiter fortgesetzt. Das Individuum wird aufgefordert, das Intervall (0.0, 0.50) in zwei gleichwahrscheinliche Teile aufzuteilen. Angenommen, das Individuum entscheidet sich für 0.42 (d.h. $p_{0.125} = 0.42$). Weiter wird das Individuum gefragt, ob es wetten würde, daß p in (0.0, 0.21) oder in (0.21, 0.42) liegt, falls man ihm sagt, daß p kleiner als 0.42 ist. Das Individuum entscheidet sich vielleicht für das letztere und nennt die Zahl 0.36, die seiner Meinung nach das Intervall (0.00, 0.42) in gleichwahrscheinliche Teile aufteilt (d.h. $p_{0.0625} = 0.36$).

Auf analoge Weise wird das Intervall (0.68, 1.00) aufgeteilt, etwa mit $p_{0.875} = 0.75$ und $p_{0.9375} = 0.80$.

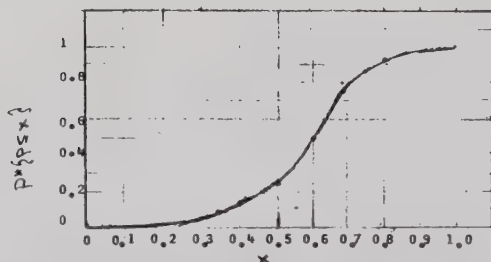
Die subjektiven Antworten des Individuum kann man in einer Tabelle zusammenfassen:

Fraktile k	subjektiver Fraktil-Wert p_k
0.0625	0.36
0.125	0.42
0.25	0.50
0.50	0.60
0.75	0.68
0.875	0.75
0.9375	0.80

Der Ausdruck $p_{0.25} = 0.50$ etwa bedeutet, daß die subjektive Wahrscheinlichkeit für das Ereignis ($p \leq 0.50$) 0.25 ist, d.h. das Individuum mißt dem Ereignis "Die Wahrscheinlichkeit dafür, daß der Angeklagte den Diebstahl begangen hat, ist 0.50 oder weniger" die Wahrscheinlichkeit 1/4 bei. Symbolisch:

$$P(p \leq 0.50) = 0.25.$$

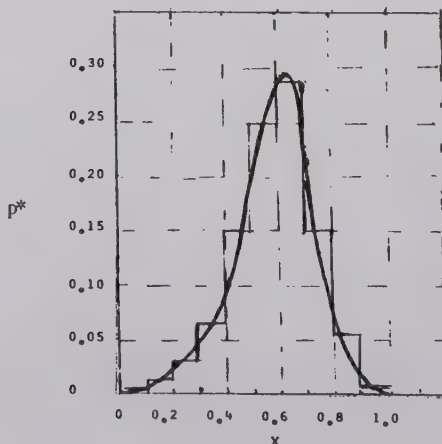
Die obige Tabelle kann durch eine graphische Darstellung veranschaulicht werden (für jeden x -Wert auf der horizontalen Achse gibt die Glättungskurve die subjektive Wahrscheinlichkeit an, daß der Wahrscheinlichkeitswert p für die obige Hypothese kleiner oder gleich x ist).



Aus der obigen Abbildung erhält man leicht die subjektive Wahrscheinlichkeit, daß p in einem bestimmten Intervall liegt. Will man z.B. den Wahrscheinlichkeitswert für das Ereignis $(0.40 < p \leq 0.50)$ bestimmen, so ist dieser wie folgt zu berechnen:

$$P^*(0.40 < p \leq 0.50) = P^*(p \leq 0.50) - P^*(p \leq 0.40) = 0.25 - 0.10 = 0.15.$$

Die folgende Abbildung zeigt ein Histogramm, das die Wahrscheinlichkeiten dafür angibt, daß p in einem der folgenden 10 Intervalle $(0, 0.10)$, $(0.10, 0.20)$, ..., $(0.90, 1.00)$ liegt. Die dabei entstehende Glättungskurve stellt eine Näherung für die sog. "subjektive Dichtefunktion" für den p -Wert dar. Der subjektive p -Wert für ein bestimmtes Intervall ist gleich der Fläche unter der Kurve innerhalb dieses Intervalls. So ist z.B. $P^*(0.40 \leq p \leq 0.60)$ die Fläche unter der Dichtekurve im Intervall $(0.40, 0.60)$ der x -Achse.



II. Die v. Neumann-Morgenstern-Axiome

Bei der Metrisierung von Nützlichkeiten und Wahrscheinlichkeiten wurde von uns oft die Hilfskonstruktion der "hypothetischen Lotterie" verwendet. Es handelt sich dabei um einen Grundbegriff der v. Neumann-Morgenstern-Nutzentheorie.

v. Neumann und Morgenstern (1947) waren die ersten, die ein Axiomensystem des "rationalen Verhaltens" aufgestellt haben, das, falls von einer Person erfüllt, die Existenz und Eindeutigkeit bis auf positiv lineare Transformationen einer Nutzenfunktion garantiert. Allerdings ist ihre Theorie in vielen praktischen Situationen (v.a. solchen des induktiven subjektiven Rasonierens) nicht anwendbar, denn es liegt ihr der objektive Wahrscheinlichkeitsbegriff zugrunde. (Inzwischen gibt es auch andere Axiomensysteme, wie z.B. dasjenige von Savage (1954) oder dasjenige von Suppes und Winet (1955)¹). Der Leser sei daran erinnert, daß es das Entscheidungsmodell von Savage war (in dem der subjektive Wahrscheinlichkeitsbegriff auftritt), das wir der rationalen juristischen Entscheidungstheorie zugrunde legten, wobei wir es durch vier zusätzliche Postulate in einer für juristische Zwecke adäquaten Weise einschränkten. Die Savage-Axiome findet der Leser im nächsten Anhang).

Grundbegriffe:

(1) Q ist eine Menge von Elementen $w, x, y, \dots$; (2) $\succsim$ ist eine binäre Relation über Q ; (3) Π ist die Menge aller reellen Zahlen $\alpha, \beta, \dots$, die streng zwischen 0 und 1 ($0 < \alpha < 1$) liegen; (4) $(\alpha, x; 1-\alpha, y)$ ist eine Operation mit $(\alpha, x; 1-\alpha, y) \in Q$ genau dann, wenn $x, y \in Q$ und $\alpha \in \Pi$.

Definitionen:

Die Elemente von Q sind Entitäten, deren Werte man feststellen will; die Elemente von Π sind Wahrscheinlichkeiten.

$x \succsim y$: y wird x vom Individuum nicht vorgezogen,

$x \succ y$: x wird y vom Individuum vorgezogen,

$x \sim y$: das Individuum ist zwischen x und y indifferent,

$(\alpha, x; 1-\alpha, y)$: Die Kombination von x mit der Wahrscheinlichkeit α und von y mit der Wahrscheinlichkeit $1-\alpha$; oder das Glücksspiel (bzw. Lotterie), das x mit der Wahrscheinlichkeit α ergibt und y mit der Wahrscheinlichkeit $1-\alpha$.

¹P. Suppes und M. Winet, An Axiomatization of Utility Based on the Notion of Utility Differences, in: 1 Management Science (1955), S. 259-270.

Die Axiome:

1. Für zwei beliebige x, y gilt nur eine der folgenden Beziehungen: $x \succ y$, $y \succ x$, $x \sim y$. (Dadurch wird die unmittelbare Vergleichbarkeit der Elemente von Q garantiert).
2. Wenn $x \succ w$ und $w \succ y$, dann $x \succ y$. (Transitivität).
3. Wenn $x \succ y$, dann $x \succ (\alpha, x; 1-\alpha, y)$ für beliebiges α . (Wenn $x \succ y$ vorgezogen wird, dann wird x auch jedem Glücksspiel über x und y vorgezogen).
4. Wenn $x \prec y$, dann $x \prec (\alpha, x; 1-\alpha, y)$ für beliebiges α . (Dual zu 3).
5. Wenn $x \succ w \succ y$, dann gibt es ein α , so daß $(\alpha, x; 1-\alpha, y) \succ w$. (D.h. wenn $x \succ w$ vorgezogen wird und dieses y , dann kann man immer ein Glücksspiel mit bestimmten Wahrscheinlichkeiten für x und y konstruieren, das w vorgezogen wird).
6. Wenn $x \prec w \prec y$, dann gibt es ein α , so daß $(\alpha, x; 1-\alpha, y) \prec w$. (Dual zu 5).
7. $(\alpha, x; 1-\alpha, y) \sim (1-\alpha, y; \alpha, x)$. (Die Reihenfolge der Alternativen in einem Glücksspiel ist irrelevant).
8. $(\alpha, (\beta, x; 1-\beta, y); 1-\alpha, y) \sim (\alpha\beta, x; 1-\alpha\beta, y)$. (Zusammengesetzte Glücksspiele können entsprechend den wahrscheinlichkeitstheoretischen Regeln zerlegt werden).
9. Wenn $x \sim y$, dann $(\alpha, x; 1-\alpha, w) \sim (\alpha, y; 1-\alpha, w)$ für beliebige α und w . (Falls y und x gleichwertig sind, dann kann man in jedem neuen Glücksspiel x durch y ersetzen).

Die obigen Axiome garantieren die Existenz einer reellwertigen Funktion f (Nutzenfunktion) über Q mit

I. $x \succ y$ genau dann, wenn $f(x) > f(y)$,

II. $f(\alpha, x; 1-\alpha, y) = \alpha f(x) + (1-\alpha)f(y)$.

Außerdem gilt:

Falls f und g zwei Funktionen über Q sind, welche die Bedingungen I und II erfüllen, dann gilt

$$g(x) = \alpha f(x) + b \quad a > 0 \quad \text{für alle } x \in Q;$$

d.h. eine (von Neumann-Morgenstern-) Nutzenfunktion ist bis auf positiv lineare Transformationen eindeutig.

III. Die Theorie von Savage (Axiomatik)

Zunächst zur Notation:

1. S ist die Menge der Zustände oder Umstände¹⁾ die Elemente von S sind $s, s', \dots$; die Teilmengen von S sind $A, B, C, \dots$; $s, s' \dots$ sind disjunkte und erschöpfende Umstände und $A, B, C, \dots$ sind Mengen von Umständen bzw. Ereignisse.
2. F ist die Menge der Konsequenzen; die Elemente von F sind $f, g, h, \dots$. (In unserer Theorie $\mathcal{R}$ mit Elementen $R_1, R_k, \dots$ oder $R_{ij}, \dots$).
3. $\mathbb{F}$ ist die Menge von Funktionen von S in F ; die Elemente von $\mathbb{F}$ sind $f, g, h, \dots$, und repräsentieren Handlungen. D.h. jede Handlung ordnet jedem Umstand eine Konsequenz zu. $f(s) = g$ bedeutet, daß die Konsequenz g dem Umstand s durch f zugeordnet ist.
4. Die binäre Relation $\preceq$ ("nicht vorgezogen")²⁾

Definitionen und Postulate:

P_1 $\preceq$ ist eine schwache Ordnung von Handlungen.

(P_1 behauptet, daß alle Paare von Handlungen vergleichbar sind).

D_1 $f \preceq g$, gegeben B , genau dann, wenn $f' \preceq g'$ für alle f' und g' , die konsistent sind jeweils mit f und g , gegeben B , und miteinander unter $\bar{B}$.

(f ist mit f' unter B konsistent, wenn $f(s) = f'(s)$ für jedes $s \in B$ (d.h. wenn f und f' jedem Umstand in B die gleiche Konsequenz zuordnen). Analog: f' und g' sind miteinander unter $\bar{B}$ konsistent genau dann, wenn $f'(s) = g'(s)$ für jedes $s \in \bar{B}$.)

P_2 Für alle f, g und B , $f \preceq g$, gegeben B , oder $g \preceq f$, gegeben B .

(P_3 behauptet, daß alle "bedingte" Entscheidungen vergleichbar sind. Für $B = S$ ist dies das Postulat P_1).

D_2 $g \preceq g'$ genau dann, wenn $f \preceq f'$, wenn $f(s) = g$, $f'(s) = g'$ für alle $s \in S$.

(d.h. Präferenzen oder Indifferenzen zwischen Konsequenzen sind definitorisch auf Präferenzen oder Indifferenzen zwischen konstanten Handlungen zurückzuführen. Eine Handlung ist konstant, wenn sie allen Umständen aus S die gleiche Konsequenz zuordnet).

D_3 B ist null oder praktisch unmöglich genau dann, wenn für alle f, g $f \preceq g$, gegeben B .

(d.h. wenn B null ist, dann sind Werte, die von Handlungen den Elementen von B zugeordnet werden, irrelevant für alle Entscheidungen).

P_3 Wenn $f(s) = g$, $f'(s) = g'$ für alle $s \in B$ und B nicht null ist, dann

$f \preceq f'$, gegeben B , genau dann, wenn $g \preceq g'$.

1) Savage selbst spricht allerdings in diesem Zusammenhang von "states of the world" (vgl. Savage, S. 8f). Im deutschen Sprachraum hat sich dafür der Ausdruck "Umstand" eingebürgert.

2) Dementsprechend sind " $>$ " als "vorgezogen" und " $\sim$ " als miteinander "äquivalent" zu lesen.

3) " $A \not\preceq B$ " ist zu lesen als: " A ist nicht wahrscheinlicher als B ".

(P3 ist eine Version des sog. "sure-thing"-Prinzips, das folgendes besagt: (1) Die Kenntnis eines Ereignisses kann Präferenzen zwischen den Konsequenzen nicht beeinflussen und (2) wenn das Ereignis nicht null ist, dann kann die Kenntnis des Ereignisses nicht Präferenzen zwischen Konsequenzen zu Indifferenzen reduzieren).

D4 $A \preceq B$ genau dann, wenn $f_A \preceq f_B$ oder $g' \preceq g$ für alle f_A, f_B, g, g' , so daß $f_A(s) = g$ für $s \in A$, $f_A(s) = g'$ für $s \in \bar{A}$, $f_B(s) = g$ für $s \in B$, und $f_B(s) = g'$ für $s \in \bar{B}$.

(d.h. "ist nicht wahrscheinlicher als" ist definiert in Ausdrücken von Präferenzen zwischen Handlungen (oder Konsequenzen). " $A \preceq B$ " hat die Interpretation: Falls das Individuum mit einer "Wahl" zwischen zwei Handlungen konfrontiert wird, von denen die erste g ergibt, wenn A sich ereignet und g' , wenn $\bar{A}$ sich ereignet, und die zweite g ergibt, wenn B und g' , wenn $\bar{B}$ sich ereignet, und es g für mindestens so wünschenswert hält wie g' und ferner die zweite Handlung "wählt", dann ist B für das Individuum mindestens so wahrscheinlich wie A . Das gilt für alle derartige g und g').

P4 Für alle A, B ist $A \preceq B$ oder $B \preceq A$.

(P4 behauptet, daß die induzierte Ordnung zwischen den Ereignissen vollständig ist).

P5 Es gibt wenigstens ein Paar von Konsequenzen ℓ, ℓ' mit $\ell' < \ell$.

(P5 stellt sicher, daß das Entscheidungsproblem nicht trivial ist).

Die Axiome P1-P5 legen fest, daß " $\preceq$ " für Ereignisse die Grundaxiome A1*-A4* der komparativen Wahrscheinlichkeit erfüllt. (Zu den Axiomen A1*-A4* vgl. S.24 dieser Arbeit).

P6 Wenn $g < h$, dann gibt es für jede Konsequenz ℓ eine Partition von S (in eine endliche Anzahl erschöpfender und disjunkter Ereignisse), so daß, wenn g' mit g und h' mit h konsistent sind, mit Ausnahme eines beliebigen Elements der Partition, mit g' und h' gleich ℓ an dieser Stelle, dann $g' < h$ oder $g < h'$ gilt.

(P6 ist das zentrale Axiom; es behauptet die Existenz eines endlich additiven P , das mit $\preceq$ konsistent ist. Savage weist darauf hin, daß P6 technisch stärker ist als nötig, um die Existenz der quantitativen subjektiven Wahrscheinlichkeit zu sichern. Allerdings meint Savage, daß es intuitiv plausibel ist; vgl. Savage, S.38f).

D5 $f \preceq g$, gegeben B ($g \preceq f$, gegeben B), genau dann, wenn $f \preceq h$, gegeben B ($h \preceq f$, gegeben B) mit $h(s) = g$ für alle s .

P7 Wenn $f \preceq g(s)$, gegeben B ($g(s) \preceq f$, gegeben B) für alle $s \in B$, dann $f \preceq g$, gegeben B ($g \preceq f$, gegeben B).

(D5 und P7 definieren und behaupten die Dominanz, die bei der Ableitung der Ordnung in $\mathbb{F}$ verlangt wird, wenn F unendlich viele Konsequenzen enthält; P7 (und D5) sind nicht notwendig, wenn F endlich ist).

Satz Seien P1-P7 erfüllt. Dann gibt es ein eindeutiges endlich-additives Wahrscheinlichkeitsmaß P auf der Potenzmenge von S mit folgenden Eigenschaften:

- (a) $A \preceq B \leftrightarrow P(A) \leq P(B)$ für alle $A, B \subseteq S$,
 (b) $(B \subseteq S, 0 \leq \rho \leq 1) \rightarrow P(C) = \rho P(B), \quad C \subseteq B$.

Außerdem gibt es eine reellwertige Funktion u auf F , für die gilt:

$$f \prec g \leftrightarrow E_u[u(f(s)), P] < E_u[u(g(s)), P] \quad \text{für alle } f, g \in \mathbb{F}.$$

Dieses u ist beschränkt und bis auf positiv lineare Transformationen eindeutig.

(Beweis: siehe z.B. bei Fishburn, Utility Theory for Decision Making (1970), S.192-210).

BIBLIOGRAPHIE

- ALEX, R., Theorie der juristischen Argumentation, Frankfurt/Main 1978.
- ARROW, K.J./KARLIN, S./SUPPES, P., Mathematical Methods in the Social Sciences, Stanford 1964.
- ARROW, K.J., Social Choice and Individual Values, New York-London-Sydney 1966.
- BALL, V.C., The Moment of Truth: Probability Theory and Standards of Proof, in: 14 Vanderbilt Law Review (1961), S.807-830.
- BALLWEG, O., Rechtswissenschaft und Jurisprudenz, Basel 1970.
- BAUER, H., Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie, Berlin 1968.
- BAYES, T., An Essay towards Solving a Problem in the Doctrine of Chances (1763). Versuch zur Lösung eines Problems der Wahrscheinlichkeitsrechnung, Leipzig 1908.
- BENDER, R. (Hrsg.), Tatsachenforschung in der Justiz, Tübingen 1972.
- BENDER, R./RÖDER, S./NACK, A., Tatsachenfeststellung vor Gericht, Bd. I, München 1981.
- BERG, W., Die verwaltungsrechtliche Entscheidung bei ungewissem Sachverhalt, Berlin 1980.
- BERGER, J.O., Statistical Decision Theory, New York 1980.
- BERKA, K./KREISER, L., Logik-Texte. Kommentierte Auswahl zur Geschichte der modernen Logik. Berlin 1971.
- BIRMINGHAM, R.L., A Model of Criminal Process: Game Theory and Law, in: 56 Cornell Law Review (1970), S.57-73.
- BIRMINGHAM, R., Remarks on 'Probability' in Law: Mostly, A Casenote and a Book Review, in: 12 Georgia Law Review (1978), S.535-551.
- BITZ, M., Entscheidungstheorie, München 1981.
- BLACKBURN, S., L.J.Cohen, The Probable and the Provable (Review), in: 44 Synthese (1980), S.149-159.
- BOCHENSKI, I.M., Die zeitgenössischen Denkmethode, München 1975.
- BOOLE, G., On the Application of the Theory of Probabilities to the Question of the Combination of Testimonies or Judgments (Originalausgabe in: Transaction of the Royal Society of Edinburgh, XXI, 1857), in: George Boole, Studies in Logic and Probability, London 1952.
- BRILMAYER, L./KORNHAUSER, L., Review: Quantitative Methods and Legal Decisions, in: 46 The Univ. of Chicago Law Review (1978), S.116-153.
- BROWN, K.S./KELLY, D.G., Playing the Percentages and the Law of Evidence, in: University of Illinois Law Forum (1970), S.23-48.
- BRUNS, R., Zivilprozeßrecht, München 1979.
- BÜHLMANN, H./LOEFFEL, H./NIEVERGELT, E., Entscheidung- und Spieltheorie, Berlin 1975.
- CARNAP, R., Logical Foundations of Probability, Chicago 1950, 1962.
- CARNAP, R., The Continuum of Inductive Methods, Chicago 1952.
- CARNAP, R./STEGMÜLLER, W., Induktive Logik und Wahrscheinlichkeit, Wien 1959.
- CARNAP, R., The aim of inductive logic, in: E.Nagel, P.Suppès and A.Tarski, Logic, Methodology and Philosophy of Science, Stanford (1962), S.303-318.
- COHEN, J./CHRISTENSEN, I., Information and choice, Edinburgh 1970.
- COHEN, L.J., The Implications of Induction, London 1970.
- COHEN, L.J., The Probable and the Provable, Oxford 1977.
- COHEN, L.J., The Logic of Proof, in: Criminal Law Review (1980), S.91-103.

- COHEN, L.J., Das Problem der Ausgangswahrscheinlichkeit bei forensischen Beweisen, in: 24 Ratio (1982), S.71-76.
- COHEN, L.J., Freedom of Proof, in: Twining, W. (Hrsg.), Facts in Law (Wiesbaden 1983), S.3-21.
- COLEMAN, R.F./WALLS, H.J., The Evaluation of Scientific Evidence, in: Criminal Law Review (1974), S.276-288.
- CORNIDES, T., Normsetzung und Entscheidungstheorie, in: 5 Rechts-theorie (1974), S.11-31.
- COWAN, T.A., Decision Theory in Law, Science and Technology, in: 17 Rutgers Law Review (1963), S.499-530.
- CULLISON, A.D., Identification by Probabilities and Trial by Arithmetic (A Lesson for Beginners in How to be Wrong with Greater Precision), in: 6 Houston Law Review (1969), S.471-518.
- CULLISON, A.D., The Model of Rules and the Logic of Decision, in: Nagel, S., Modeling the Criminal Justice System (Beverly Hills 1977), S.225-246.
- DÖRING, E., Die Erforschung des Sachverhalts im Prozeß, Berlin 1964.
- DUBS, H., Analytische Bewertung als Grundlage richterlicher Strafzumessung, in: Festgabe zum schweizerischen Juristentag, Basel 1963.
- EDWARDS, A.F., Likelihood. An Account of the Statistical Concept of Likelihood and its Application to Scientific Inference, Cambridge 1972.
- EDWARDS, W., The Theory of Decision Making, in: 51 Psychological Bulletin (1954), S.380-415.
- EDWARDS, W./LINDMANN, H./SAVAGE, L.J., Bayesian Statistical Inference for Psychological Research, in: 70 Psychological Review (1963), S.193-242.
- EDWARDS, W./TVERSKY, A., Decision Making, Victoria 1967.
- EGGLESTON, R., Probabilities and Proof, in: 4 Melbourne Univ. Law Review (1963), S.180-211.
- EGGLESTON, R., Beyond Reasonable Doubt, in: 4 Monash Univ. Law Review (1977), S.1-22.
- EGGLESTON, R., Evidence, Proof and Probability, London 1978.
- EGGLESTON, R., Subjective Probability and the Law, in: Bell, C.R. (Hrsg.), Uncertain Outcomes (Lancaster 1979), S.127-145.
- EGGLESTON, R., The Probability Debate, in: Criminal Law Review (1980), S.678-688.
- EGGLESTON, R./Lindley, D.V., The Problem of Missing Evidence, in: 99 The Law Quarterly Review (1983), S.86-99.
- EKELÖF, P.O., Free Evaluation of Evidence, in: 8 Scandinavian Studies in Law (1964), S.45-66.
- ENGISCH, K., Wahrheit und Richtigkeit im juristischen Denken, München 1963.
- ESSEN-MÖLLER, E., Die Beweiskraft der Ähnlichkeit im Vaterschaftsnachweis, in: Mitteilungen der Anthropologischen Gesellschaft in Wien, Bd. LXVIII (1938), S.9-53.
- ESSER, J., Vorverständnis und Methodenwahl in der Rechtsfindung, Frankfurt/Main, 1972.
- EVERS, A., Begriff und Bedeutung der Wahrscheinlichkeit für die richterliche Beweismwürdigung, Freiburg 1979.
- FAIRLEY, W.B./MOSTELLER, F., A Conversation about Collins, in: 41 Univ. of Chicago Law Review (1974), S.242-253.
- FAIRLEY, W.B., Probabilistic Analysis of Identification Evidence, in: Wendt/Vlek (Hrsg.), Utility, Probability, and Human Decision Making (1975), S.233-256.
- FERSCHEL, F., Nutzen- und Entscheidungstheorie, Opladen 1975.
- FINE, T.L., Theories of Probability: An Examination of Foundations, New York - London 1973.

- de FINETTI, B., La Prévision: Ses Lois Logiques, Ses Sources Subjectives, in: Annales de l'Institut Henri Poincaré, Bd.7 (1937), S.1-68.
- FINKELSTEIN, M.O./FAIRLEY, W.B., A Bayesian Approach to Identification Evidence, in: 83 Harvard Law Review (1970), S.489-517.
- FINKELSTEIN, M.O./FAIRLEY, W.B., A Comment on "Trial by Mathematics", in: 84 Harvard Law Review (1971), S.1801-1809.
- FINKELSTEIN, M.O., Quantitative Methods in Law: Studies in the Application of Mathematical Probability and Statistics to Legal Problems, New York 1978.
- FINNEY, D.J., Probabilities Based on Circumstantial Evidence, in: 72 J. of the American Stat. Ass. (1977), S.316-318.
- FISHBURN, P.C., Decision and Value Theory, New York 1964.
- FISHBURN, P.C., Utility Theory, in: 14 Management Science (1968), S.335-378.
- FISHBURN, P.C., Utility Theory for Decision Making, New York 1970.
- FISHBURN, P.C., Mathematics of Decision Theory, Mouton-The Hague-Paris 1970.
- FISHBURN, P.C., The Theory of Representative Majority Decision, in: 39 Econometrica (1971), S.273-284.
- FRANZEN, W., Zur Redundanztheorie der Wahrheit. Ein historischer und systematischer Überblick, in: 12 Conceptus (1978), S.54-69; (1979) S.47-62.
- FRANZEN, W., Zur neueren Wahrheitsdiskussion: Redundanztheorie versus Korrespondenztheorie der Wahrheit, in: 35 Zeitschrift für philosophische Forschung (1981), S.73-90.
- GÄFGEN, G., Theorie der wirtschaftlichen Entscheidung, Tübingen 1974.
- GÄFGEN, G., Formale Theorie des politischen Handelns: Wissenschaftliche Politik als rationale Wahl von Strategien, in: Maier/Ritter/Matz (Hrsg.), Politik und Wissenschaft (München 1971), S.209-257.
- GÄRDENFORS, P., On the Logic of Relevance, in: 37 Synthese (1978), S.351-367.
- GELFAND, A.E./SOLOMON, H., Analyzing the Decision-Making Process of the American Jury, in: 70 J. of the American Stat. Ass. (1975), S.305-310.
- GERJUOY, E., The Relevance of Probability Theory to Problems of Relevance, in: 18 Jurimetrics J. (1977), S.1-39.
- GOLDING, M.P., Principled Decision-Making and the Supreme Court, in: 63 Columbia Law Review (1963), S.35-58.
- GOOD, I., Weight of Evidence, Corroboration, Explanatory Power, Information and the Utility of Experiments, in: 22 J. of the Royal Stat. Society (1960), S.319-331.
- GOOD, I., Probability and the Weighing of Evidence, London 1950.
- GOTTINGER, H.W., Subjektive Wahrscheinlichkeiten, Göttingen 1974.
- GOTTINGER, H.W., Grundlagen der Entscheidungstheorie, München 1974.
- GOTTWALD, P., Schadenszurechnung und Schadensschätzung, München 1979.
- GRASSBERGER, R., Die Strafzumessung, Wien 1932.
- GREGER, R., Beweis und Wahrscheinlichkeit, Köln-Berlin-Bonn-München 1978.
- HAAG, K., Rationale Strafzumessung. Ein entscheidungstheoretisches Modell, Köln-Berlin-Bonn-München 1970.
- HABERMAS, J., Wahrheitstheorien, in: Fahrenbach (Hrsg.), Wirklichkeit und Reflexion (Festschrift für W.Schulz) (1973), S.221-265.
- HACKING, I., Logic of Statistical Inference, Cambridge 1965.
- HALLDÉN, S., Preference Logic and Theory Choice, in: 16 Synthese (1966), S.307-320.

- HARSANYI, J.C., Bayesian Decision Theory and Utilitarian Ethics, in: 68 American Economic Association (1978), S.223-228.
- HARSANYI, J.C., A new Solution Concept for both Cooperative and Non-cooperative Games, Opladen 1979.
- HAVERY, R.O., Baconian and Pascalian Probability in Forensic Proof, in: Crim. Law Review (1980), S.743-747.
- HEMPEL, C.G., On the Logical Positivists' Theory of Truth, in: 2 Analysis (1934/35), S.49-59.
- von HIPPEL, R., Gefahrenurteile und Prognosenentscheidungen in der Strafrechtspraxis, Berlin 1972.
- IVERSEN, G.R., Operationalizing the Concept of Probability in Legal-Social Science Research, in: 5 Law and Society Review (1971), S.331-333.
- JACKSON, J.D., Probability and Mathematics in Court Fact-Finding, in: 31 Northern Ireland Legal Quarterly (1980).
- JAMES, G.F., Relevancy, Probability and the Law, in: 29 California Law Review (1941), S.689-705.
- JEFFREY, R.C., Logik der Entscheidungen, dt. Übers. von: The Logic of Decision, Wien-München 1967.
- KADANE, J.B., Juries Hearing Death Penalty Cases: Statistical Analysis of a Legal Procedure, in: 78 J. of the American Stat. Ass. (1983), S.544-552.
- KAHNEMANN, D./TVERSKY, A., On the Interpretation of Intuitive Probability: A Reply to Jonathan Cohen, in: 7 Cognition (1979), S.409-411.
- KAPLAN, J., Decision Theory and the Factfinding Process, in: 20 Stanford Law Review (1968), S.1065-1092.
- KAYE, D., The Laws of Probability and the Law of the Land, in: 47 The Univ. of Chicago Law Rev. (1979), S.34-56.
- KAYE, D., Probability Theory Meets Res Ipsa Loquitur, in: 77 Michigan Law Review (1979), S.1456-1484.
- KAYE, D., Naked Statistical Evidence, in: 89 The Yale Law Review (1980), S.601-611.
- KEGEL, G., Der Individualanscheinsbeweis und die Verteilung der Beweislast nach überwiegender Wahrscheinlichkeit, in: Festgabe für H. Kronstein (Karlsruhe 1967), S.321-344.
- KEYNES, J.M., A Treatise on Probability, London 1957 (1921).
- KILIAN, W., Juristische Entscheidung und elektronische Datenverarbeitung, Frankfurt 1974.
- KIRSCH, W., Entscheidungsprozesse, I-III, Wiesbaden 1970.
- KINGSTON, C.R./KIRK, P.L., The Use of Statistics in Criminalistics, in: 55 J. of Criminal Law, Criminology and Police Science (1964), S.514-521.
- KINGSTON, C.R., Applications of Probability Theory in Criminalistics, in: 60 J. of the American Stat. Ass. (1965), S.70-80, S.1028-1034.
- KINGSTON, C.R., Probability and Legal Proceedings, in: 57 J. of Criminal Law, Criminology and Police Science (1966), S.93-98.
- KOCH, H.J./RÜßMANN, H., Juristische Begründungslehre, München 1982.
- KORNSTEIN, D.J., A Bayesian Model of Harmless Error, in: 2 The Journal of Legal Studies (1976), S.121-145.
- KRANTZ, D.H./LUCE, R.D./SUPPES, P./TVERSKY, A., Foundations of Measurement, Vol. I, New-York, London 1971.
- KRELLE, W., Präferenz- und Entscheidungstheorie, Tübingen 1968.
- von KRIES, J., Über die Begriffe der Wahrscheinlichkeit und Möglichkeit und ihre Bedeutung im Strafrechte, in: 9 ZStW (1889), S.528-537.

- KRITZ, J., Subjektive Wahrscheinlichkeiten und Entscheidungen, Frankfurt/Main 1982.
- KÜNZE, H. P./KRELLE, W., Einführung in die mathematische Optimierung, Zürich 1969.
- von KUTSCHERA, F., Wissenschaftstheorie, I, II, München 1972.
- von KUTSCHERA, F., Einführung in die Logik der Normen, Werte und Entscheidungen, Freiburg 1973.
- von KUTSCHERA, F., Der subjektive Wahrscheinlichkeitsbegriff, in: 20 Synthese (1969), S. 84-103.
- KYBURG, H. E./SMOKLER, H. E., Studies in subjective Probabilities, New York 1964.
- LASSWELL, H. D./McDOUGAL, M. S., Jurisprudence in Policy-Oriented Perspective, in: 19 Univ. of Florida Law Review (1966/67), S. 486-513.
- LEMPERT, R. O., Modeling Relevance, in: 75 Michigan Law Review (1977), S. 1021-1057.
- LEVI, I., Recent Work in Probability and Induction (Review), in: 16 Synthese (1966), S. 234-244.
- LEVI, I., Are Statistic Hypotheses Covering Laws?, in: 20 Synthese (1969), S. 297-307.
- LEVI, I., Support and Surprise: L. J. Cohen's View of Inductive Probability, in: 30 Brit. J. Phil. Sci. (1979), S. 279-292.
- LIDDLE, T. H., Mathematical and Statistical Probability As a Test of Circumstantial Evidence, in: 19 Case Western Law Rev. (1968), S. 254-278.
- LINDLEY, D. V., Probabilities and the Law, in: Wendt/Vlek (Hrsg.), Utility, Probability and Human Decision Making (1975), S. 223-232.
- LUCE, R. D., A Probabilistic Theory of Utility, in: 26 Econometrica (1958), S. 193-224.
- LUCE, R. D., Sufficient conditions for the existence of a finitely additive probability measure, in: 38 Ann. Math. Statist. (1967), S. 780-786.
- LUCE, R. D./RAIFFA, H., Games and Decisions, New York 1957.
- LUCE, R. D./KRANTZ, D. H., Conditional Expected Utility, in: 39 Econometrica (1971), S. 253-271.
- MAASSEN, B. U., Beweisprobleme im Schadensersatzprozeß, Köln-Berlin-München 1975.
- MAIER, H./RITTER, K./MATZ, U., Politik und Wissenschaft, München 1971.
- MAKKONEN, K., Zur Problematik der juridischen Entscheidung, Turku 1965.
- MARSCHAK, J., Intersubjektive Wahrscheinlichkeit, in: Heidelberger Jahrbücher XVII, 1973, S. 14-26.
- MARSHALL, C. R./WISE, J. A., Juror Decisions and the Determination of Guilt in Capital Punishment Cases: A Bayesian Perspective, in: Wendt/Vlek (Hrsg.), Utility, Probability, and Human Decision Making (1975), S. 257-269.
- MARSHALL, J., Evidence, Psychology, and the Trial: Some Challenges to Law, in: 63 Columbia Law Review (1963), S. 197-231.
- McBAINE, J. P., Burden of Proof: Degrees of Belief, in: 32 California Law Review (1944), S. 242-268.
- MENGES, G. (Hrsg.), unter Mitarbeit von B. Leitner, Bibliographie zur statistischen Entscheidungstheorie 1950-1967, Köln-Opladen 1968.
- MENGES, G., Grundmodelle wirtschaftlicher Entscheidungen, Düsseldorf 1974.
- MILANICH, P. G., Decision Theory and Standards of Proof, in: 5 Law and Human Behavior (1981), S. 87-96.
- MODE, E. B., Probability and Criminalistics, in: 58 J. of the American Stat. Ass. (1963), S. 628-640.

- MOTSCH, R., Vom Prozeß als Beweis zum Überwiegensprinzip, in: Gedächtnisschrift für Jürgen Rödig, Berlin-Heidelberg-New York 1978, S. 334-348.
- MOTSCH, R., Vom rechtsgenügenden Beweis. Zur Entscheidung von Zivilsachen nach Wahrscheinlichkeit unter besonderer Berücksichtigung der Abstammungsfeststellung, Berlin 1983.
- MUSIELAK, H.-J., Die Grundlagen der Beweislast im Zivilprozeß, Berlin-New York 1975.
- MUSIELAK, H.-J., Das Överviktsprinzip. Zum Verhältnis von richterlicher Überzeugung und Wahrscheinlichkeit, in: Festschrift für Gerhard Kegel, Frankfurt 1977, S. 451-471.
- NAGEL, S.S., Modeling the Criminal Justice System, Beverly Hills 1977.
- NAGEL, S.S./NEEF, M.G., The Legal Process Modeling the System, Beverly Hills 1977.
- NAGEL, S.S./NEEF, M.G., Decision Theory and the Legal Process, Toronto 1979.
- NELL, E.L., Wahrscheinlichkeitsurteile in juristischen Entscheidungen, Berlin-München 1983.
- NESSON, C.R., Reasonable Doubt and Permissive Inferences: The Value of Complexity, in: 92 Harvard Law Review (1979), S. 1187-1225.
- von NEUMANN, J./MORGENSTERN, O., Theory of Games and Economic Behavior, Princeton (1944); dt. Übers.: Spieltheorie und wirtschaftliches Verhalten, Würzburg 1967.
- NEURATH, O., Soziologie im Physikalismus, in: 2 Erkenntnis (1931), S. 393-431.
- NEURATH, O., Protokollsätze, in: 3 Erkenntnis (1932/1933), S. 204-214, 346-362.
- NISSON, K.D., Theorie der Begutachtung, in: 172 Archiv für Kriminologie (1983), S. 143-152.
- OCKELTON, C.M.G., Mathematische Wahrscheinlichkeiten und die Bewertung bestätigender oder konvergierender Beweismittel, in: 24 Ratio (1982), S. 61-70.
- OSTERBURG, J.W./PATHASARATHY, T./RAGHAVAN, T.E.S./SCLOVE, S.L., Development of a Mathematical Formula for the Calculation of Fingerprint Probabilities Based on Individual Characteristics, in: 72 J. of the American Stat. Ass. (1977), S. 772-778.
- OWEN, G., Spieltheorie, Berlin 1971.
- PARKER, J.B., A Statistical Treatment of Identification Problems, in: 6 J. of Forensic Scie. Soc. (1966), S. 33-39.
- PARKER, J.B., The Mathematical Evaluation of Numerical Evidence, in: 6 J. of Forensic Scie. Soc. (1966), S. 134-144.
- PENROD, S./HASTIE, R., Models of Jury Decision Making: A Critical Review, in: 86 Psychological Bulletin (1979), S. 462-492.
- PODLECH, A., Rechnen und Entscheiden. Mathematische Modelle juristischen Argumentierens, Berlin 1977.
- POPPER, K., Logik der Forschung, Wien 1934 (Tübingen ⁶1976).
- PRATT, J.W./RAIFFA, H./SCHLAIFFER, R., Introduction to the Statistical Decision Theory, New York 1965.
- PRIESTER, J.H., Rationalität und Normkritik, Saarbrücken 1972.
- PUNTEL, L.B., Wahrheitstheorien in der neueren Philosophie, Darmstadt 1978.
- RAIFFA, H., Einführung in die Entscheidungstheorie, München 1973; dt. Übers. von: Decision Analysis, Reading, Mass. 1968.
- RAMSEY, F.P., Facts and Propositions, in: Aristotelian Soc. Supplementary, Vol. VII (1927), S. 153-170.
- RAMSEY, F.P., The Foundations of Mathematics, and Other Logical Essays, London ²1978 (1931).

- RAPOPORT, A., Der spieltheoretische Ansatz zur Entscheidungstheorie, Bochum 1982.
- RAWLS, J., Eine Theorie der Gerechtigkeit, Frankfurt 1975; dt. Übers. von: A Theory of Justice (1971).
- REICHENBACH, H., The Direction of Time, Berkely 1956.
- RESCHER, N., The Coherence Theory of Truth, Oxford 1973.
- RÖDIG, J., Die Theorie der gerichtlichen Erkenntnisverfahren, Berlin u. 1973.
- SAVAGE, L. J., The Theory of Statistical Decision, in: 46 J. of the American Stat. Ass. (1951), S. 55-67.
- SAVAGE, L. J., The Foundations of Statistics, New York (1954), 1972.
- SCHLICK, M., Über das Fundament der Erkenntnis, in: 4 Erkenntnis (1934) S. 79-99.
- SCHLINK, B., Inwieweit sind juristische Entscheidungen mit entscheidungstheoretischen Modellen zu erfassen und praktisch zu bewältigen?, in: Rechtstheorie als Grundlagenwissenschaft der Rechtswissenschaft. Jahrbuch für Rechtssoziologie und Rechtstheorie, Bd. II, Düsseldorf 1972, S. 322-344.
- SCHMIDT, J., Noch einmal: Wahrheitsbegriff und Rechtswissenschaft, in: JUS (1973), S. 202-207.
- SCHNEEWEIS, H., Entscheidungskriterien bei Risiko, Berlin-Heidelberg-New York 1967.
- SCHNEIDER, E., Beweis und Beweiswürdigung, unter besonderer Berücksichtigung des Zivilprozesses, München 1978.
- SCHREIBER, R., Theorie des Beweiswertes für Beweismittel im Zivilprozeß Berlin-Heidelberg-New York 1968.
- SCHUM, D. A., A Review of A Case Against Blaise Pascal and his Heirs, in: 77 Michigan Law Review (1979), S. 446-483.
- SEIFERT, H. G., Spieltheorie. Überblick über neuere Entwicklungen, Regensburg 1980.
- SELTEN, E., Anwendung der Spieltheorie auf die politische Wissenschaft, in: Maier/Ritter/Matz (Hrsg.), Politik und Wissenschaft (München 1971), S. 287-320.
- SEN, A. K., Collective Choice and Social Welfare, London 1970.
- SHAFER, G., A Mathematical Theory of Evidence, Princeton 1976.
- SHUBIK, M., Spieltheorie und Sozialwissenschaften, Frankfurt 1965.
- SIGAL, P., Judicial Use, Misuse, and Abuse of Statistical Evidence, in: 47 J. of Urban Law (1969), S. 165-190.
- SIMON, R., "Beyond a Reasonable Doubt" - An Experimental Attempt at Quantification, in: 6 J. of Applied Behavioral Science (1970), S. 203-209.
- SIMON, R./MAHAN, L., Quantifying Burdens of Proof. A View from the Bench, the Jury, and the Classroom, in: 5 Law and Society Review (1971), S. 319-330.
- SINNREICH, J., (Hrsg.), Zur Philosophie der idealen Sprache, München 1972.
- SKIRBEKK, G., Wahrheitstheorien, Frankfurt/Main 1977.
- SMITH, R. L./CHARROW, R. P., Upper and Lower Bounds for Probability of Guilt Based on Circumstantial Evidence, in: 70 J. of the American Stat. Ass. (1975), S. 555-560.
- SPENDEL, G., Wahrheitsfindung im Strafprozeß, in: JUS (1964), S. 465-473.
- SPOHN, W., Grundlagen der Entscheidungstheorie, München 1978.
- STEGMÜLLER, W., Das Wahrheitsproblem und die Idee der Semantik, Wien 1957.
- STEGMÜLLER, W., Hauptströmungen der Gegenwartsphilosophie, Bd. I, Stuttgart 4 Aufl. 1969.

- STEGMÜLLER, W., Das Problem der Induktion: Humes Herausforderung und moderne Antworten, in: Lenk, H. (Hrsg.), Neue Aspekte der Wissenschaftstheorie, Braunschweig 1971, S.13-74.
- STEGMÜLLER, W., Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie, Bd.I, Wissenschaftliche Erklärung und Begründung, Berlin-Heidelberg-New York 1974.
- STEGMÜLLER, W., Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie, Bd.IV, Personelle und Statistische Wahrscheinlichkeit, Berlin u.a. 1973.
- STOEBUCK, W.B., Relevancy and the Theory of Probability, in: 51 Iowa Law Review (1966), S.849-861.
- STONE, J., Burden of Proof and the Judicial Process, in: 60 The Law Quarterly Review (1944), S.262-284.
- STRIPINIS, D., Probability Theory and Circumstantial Evidence: Implications from a Mathematical Analysis, in: 22 Jurimetrics J. (1981), S.59-82.
- SUPPES, P./WINET, M., An Axiomatization of Utility Based on the Notion of Utility Differences, in: 1 Management Science (1955), S.259-270.
- SUPPES, P./ZINNES, J.L., Basic Measurement Theory, in: Luce, R.D./Bush, R.R., Galanter, E. (Hrsg.), Mathematical Psychology, New York - London 1963, S.1-76.
- TARSKI, A., Der Wahrheitsbegriff in den formalisierten Sprachen. Deutsche Originalausgabe in: Studia Philosophica Commentarii Societatis Philosophicae Polonorum, Vol.I, Leopolis (Lemberg) 1935. Abgedruckt in: Berka-Kreiser, Logik-Texte (Berlin 1971), S.447-559.
- TARSKI, A., Die semantische Konzeption der Wahrheit und die Grundlagen der Semantik, in: Sinnreich, J. (Hrsg.), Zur Philosophie der idealen Sprache (1972). Abgedruckt in: Skirbekk, G., Wahrheitstheorien (Frankfurt/Main 1977), S.140-188.
- TRIBE, L.H., Trial by Mathematics: Precision and Ritual in the Legal Process, in: 84 Harvard Law Review (1971), S.1329-1393.
- TRIBE, L.H., A Further Critique of Mathematical Proof, in: 84 Harvard Law Review (1971), S.1810-1820.
- TURKOWSKI, C., Über die Wahrscheinlichkeit des Auftretens gleicher Muster, in: 151 Archiv für Kriminologie (1973), S.144-148.
- TVERSKY, A./KAHNEMAN, D., On the Interpretation of Intuitive Probability: A Reply to Jonathan Cohen, in: 7 Cognition (1979), S.409-411.
- TWINNING, W. (Hrsg.), Facts in Law, Wiesbaden 1983.
- WÄLDE, T., Entscheidungstheoretische Perspektiven für die Rechtsanwendung, in: 6 Rechtstheorie (1975), S.205-246.
- WALLS, H.J., What is "Reasonable Doubt?" A Forensic Scientist Looks at the Law, in: Crim. Law Review (1971), S.458-470.
- WALTER, G., Freie Beweiswürdigung. Eine Untersuchung zu Bedeutung, Bedingungen und Grenzen der freien richterlichen Überzeugung, Tübingen 1979.
- WASSERSTROM, R.A., The Judicial Decision, Stanford-London 1961.
- WECHSLER, H., Principles, Politics, and Fundamental Law, Cambridge/Mass. 1961.
- WEITNAUER, H., Wahrscheinlichkeit und Tatsachenfeststellung, in: Karlsruher Forum 1966 (Versicherungsrecht 1966, Beiheft), S.3-20.
- WENDT, D./VLEK, C. (Hrsg.), Utility, Probability, and Human Decision Making, Dordrecht-Boston 1975.
- WILLIAMS, G., The Mathematics of Proof, I, II, in: Crim. Law Review (1979), S.297-308, S.340-354.
- WILLIAMS, G., A Short Rejoinder, in: Crim. Law Review (1980), S.103-107.

Liste der Symbole:

Logische Symbole

" $\wedge$ " für "und" (Konjunktion), " $\vee$ " für "oder" (Adjunktion), " $\neg$ " für "nicht" (Negation), " $\rightarrow$ " für "wenn...,dann..." (Konditional), " $\leftrightarrow$ " für "genau dann..., wenn..." (Bikonditional), " $\forall$ " für "für alle gilt..." (Allquantifikation), " $\exists$ " für "es gibt..." (Existenzquantifikation);

" $\vdash$ " für "impliziert" (logische Folgerung).

Mengentheoretische Symbole

" $\in$ " für "Element von" (Elementschaftsrelation), " $\{ \}$ " für "Menge";

die den logischen Zeichen " $\wedge$ ", " $\vee$ ", " $\neg$ " entsprechenden mengentheoretischen Symbole:

" $\cap$ " für Durchschnitt, " $\cup$ " für Vereinigung, " $\neg$ " (über einem Mengensymbol) für Komplement;

" $\subseteq$ " für "Teilmenge von...", " $\setminus$ " für "ohne" (Differenz, eine andere Bezeichnung: " $-$ ").

Grundsymbole

A,B,C,D,E,... je nach Kontext entweder beliebige Ereignisse oder beliebige Sachverhalte (Tatbestände, Sätze, Propositionen...)

P: Wahrscheinlichkeitsfunktion, P(): Wahrscheinlichkeit,

u: Nutzenfunktion (Nutzen), E_u , $E_u(\cdot)$: Erwartungswert des Nutzens;

" $+$ ": Addition, " Σ ": Summe, " ∞ ": Unendlichkeitszeichen, " Π ": Produkt,

" $\cdot$ ": Multiplikation;

" $<$ ": Präferenzrelation, " $\sim$ ": Indifferenzrelation;

" $A \preceq B$ ": je nach Kontext: "B ist mindestens ebenso wahrscheinlich wie A",

"B ist mindestens so wünschenswert wie A", "A wird B nicht vorgezogen" etc.

" $A < B$ ": je nach Kontext: "B ist wahrscheinlicher, wünschenswerter etc. als A",

" $A \approx B$ ": je nach Kontext: "A und B sind gleichermaßen wahrscheinlich, wünschenswert etc." (entsprechend: " $A \succcurlyeq B$ ", " $A \preccurlyeq B$ ").

" $\leq$ ": "Kleiner-als-Relation" zwischen reellen Zahlen, " $=$ ": Gleichheitsrelation zwischen reellen Zahlen (entsprechend " $\geq$ ", " $<$ ", " $>$ ").

" $\approx$ " für "ungefähr" (äquivalent dazu gelegentlich " $\sim$ " über der am meisten rechts stehenden reellen Zahl).

Sind Argumente einer Wahrscheinlichkeitsfunktion Ereignisse (Mengen), so werden mengentheoretische Symbole verwendet; sind sie dagegen Sachverhalte, dann logische Zeichen: P(A) ist die Wahrscheinlichkeit für die Negation des Ereignisses A, P($\neg$ A) ist eine solche für die Negation des Sachverhalts A. Im Text werden die beiden Schreibweisen äquivalent verwendet.

Lebenslauf:

Eleonora Bourmistrov-Jüttner (geb. Bourmistrov), geboren am 3.7.1955 in Moskau/UdSSR. Besuch einer Oberschule in Moskau, Abitur 1972. 1973 Übersiedlung zur Mutter nach München; deutsche Staatsbürgerschaft seit 1981. Ab WS 1975 Studium der Philosophie als Hauptfach an der Ludwig-Maximilians-Universität in München; ab WS 1979 Studium der Wissenschaftstheorie und Logik als Hauptfach; Magisterprüfung 1982. Abschließend Studium der Mathematik in München. Doktorprüfung in Philosophie, Wissenschaftstheorie und Logik im SS 1986.

